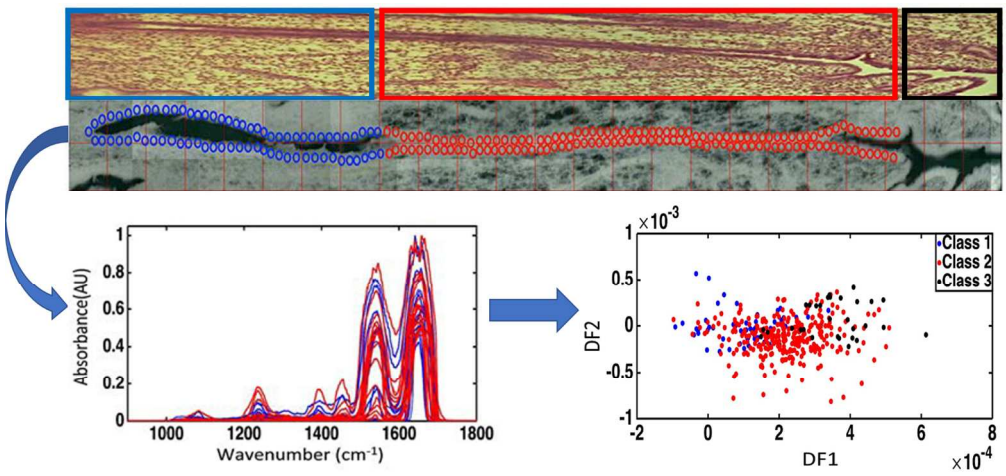
Synchrotron- and focal plane array-based Fourier-transform infrared spectroscopy differentiates the Basalis and Functionalis epithelial endometrial regions and identifies putative stem cell regions of human endometrial glands

Journal:	*Analytical and Bioanalytical Chemistry*
Manuscript ID	ABC-00221-2018.R2
Type of Paper:	Research Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Theophilou, Georgios; Lancashire Teaching Hospitals NHS Trust Morais, Camilo; University of Central Lancashire, School of Pharmacy and Biomedical Sciences Halliwell, Diane; University of Central Lancashire, School of Pharmacy and Biomedical Sciences Lima, Kassio; Universidade Federal do Rio Grande do Norte, Chemistry Drury, Josephine; University of Liverpool Institute of Translational Medicine, Women's and Children's Health Martin-Hirsch, Pierre; Lancashire Teaching Hospitals NHS Trust, Stringfellow, Helen; Lancashire Teaching Hospitals NHS Trust Hapangama, Dharani; University of Liverpool Institute of Translational Medicine, Women's and Children's Health Martin, Francis; University of Central Lancashire, School of Pharmacy and Biomedical Sciences
Keywords:	Biospectroscopy, Cell lineage, Chemometrics, Endometrium, FTIR spectroscopy, Stem cells



Graphical abstract

354x166mm (120 x 120 DPI)

Synchrotron- and focal plane array-based Fourier-transform infrared spectroscopy differentiates the Basalis and Functionalis epithelial endometrial regions and identifies putative stem cell regions of human endometrial glands

Georgios Theophilou[1], Camilo L. M. Morais[2], Diane E. Halliwell[2], Kássio M. G. Lima[3], Josephine Drury[4], Pierre L. Martin-Hirsch[5], Helen F. Stringfellow[5], Dharani K. Hapangama[4], Francis L. Martin[2*]

[1]Department of Gynaecology, Leeds Teaching Hospitals NHS Foundation Trust, Leeds, United Kingdom; [2]School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston PR1 2HE, United Kingdom; [3]Biological Chemistry and Chemometrics, Institute of Chemistry, Federal University of Rio Grande do Norte, Natal 59072-970, Brazil; [4]Department of Obstetrics and Gynaecology, Liverpool Women's NHS Foundation Trust, Liverpool, United Kingdom; [5]Department of Obstetrics and Gynaecology, Lancashire Teaching Hospitals NHS Foundation Trust, Preston, United Kingdom

***Corresponding author:** Francis L. Martin, School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston PR1 2HE, United Kingdom. Email: flmartin@uclan.ac.uk; Tel.: +44 01772 896482

Abstract

The cyclical process of regeneration of the endometrium suggests that it may contain a cell population that can provide daughter cells with high proliferative potential. These cell lineages are clinically significant as they may represent clonogenic cells that may also be involved in tumourigenesis as well as endometriotic lesion development. To determine whether the putative stem cell location within human uterine tissue can be derived using vibrational spectroscopy techniques, normal endometrial tissue was interrogated by two spectroscopic techniques. Paraffin-embedded uterine tissues containing endometrial glands were sectioned to 10-µm-thick parallel tissue sections and were floated onto BaF_2 slides for synchrotron radiation-based Fourier-transform infrared (SR-FTIR) microspectroscopy and globar Focal Plane Array-based FTIR spectroscopy. Different spectral characteristics were identified depending on the location of the glands examined. The resulting infrared spectra were subjected to multivariate analysis to determine associated biophysical differences along the length of longitudinal and crosscut gland sections. Comparison of the epithelial cellular layer of transverse gland sections revealed alterations indicating the presence of putative transient amplifying-like cells in the basalis and mitotic cells in the functionalis. SR-FTIR microspectroscopy of the base of the endometrial glands identified the location where putative stem cells may reside at the same time pointing towards $\nu_s PO_2^-$ in DNA and RNA, nucleic acids and Amide I and II vibrations as major discriminating factors. This study supports the view that vibration spectroscopy technologies are a powerful adjunct to our understanding of the stem cell biology of endometrial tissue.

Keywords: Biospectroscopy; cell lineage; chemometrics; endometrium; FTIR spectroscopy; stem cells

Introduction

The endometrium is a highly regenerative tissue that undergoes monthly cycles of shedding and regeneration under the influence of oestrogen and progesterone [1, 2]. It comprises two regions, the functionalis and the basalis. The functionalis is the region that is shed with menstruation and is composed of glands lined by pseudostratified columnar epithelium during the proliferative phase of the menstrual cycle, and columnar epithelium in the secretory phase. It is surrounded by vascularised stroma. The basalis comprises of the bases of the glands, surrounded by vascular endothelium and denser stroma [3]. The basalis layer extends into the myometrium. The endometrial/myometrial junction is the transition zone between the glandular epithelium and the stroma of the endometrium and the inner myometrium, which are not separated by a submucosal layer. Despite this, they are clearly distinguishable by light microscopy. Embryologically, these areas are of similar origin arising from the paramesonephric ducts, while the outer myometrium is thought to be of non-paramesonephric origin.

Adult stem cells are rare undifferentiated cells that function to maintain tissue homeostasis by generating replacement cells for tissues during routine cellular turnover or for the repair of tissues during injury [4]. Their defining properties are proliferation and self-renewal and differentiation into multiple lineages depending on the tissues they aim to regenerate [5]. They reside amongst niche cells and extracellular matrix which transmit signals that regulate their activity, at the same time guarding their genetic stability [4]. In the endometrium there are three types of stem cells including epithelial, endothelial and mesenchymal stem cells [6]. It has been hypothesised that stem cells reside within the epithelial layer lining the glands of the

basalis layer as well as the surrounding stroma and are responsible for the cyclical regeneration of the functionalis layer [1]. At the present time, there are no conclusive markers to isolate epithelial progenitor cells in menstrual blood while stromal progenitor cells have been identified [7]. This suggests that the former may reside exclusively in the luminal epithelium of the basalis layer [8]. Their daughter cells are termed transient-amplifying cells and migrate to the functionalis layer resulting in its proliferation. Several techniques have been exploited to confirm the presence, location and activity of endometrial stem cells. These include cloning studies [9, 10], immunochemistry studies [11, 12] and regeneration studies [13]. Expression of immunochemical markers for such stem cells was demonstrated in human endometrium but the exact location of most of these cells has not been pinpointed [14]. The SSEA1 expressing epithelial cells from basalis glands have shown some progenitor activity *in vitro* [12], but their abundance suggests that only a subpopulation of SSEA1 expressing epithelial cells from basalis glands have adult stem cell properties [12].

Stem cells may be implicated in uterine carcinogenesis. Cancer stem cells possess similar properties to stem cells in terms of differentiation and self-renewal potential. They differ from benign adult stem cells in that their growth potential is no longer controlled by signals from the surrounding niche cells therefore they proliferate uncontrollably and differentiate into unpredictable cellular lineages [15]. Cancer stem cells may be derived from resident adult stem cells through genetic or epigenetic changes [15].

Endometrial stem cells may also be implicated in the pathogenesis of endometriosis. Sampson's retrograde menstruation theory states that endometrial fragments that enter the peritoneal cavity through the fallopian tubes during

menstruation, implant into peritoneal surfaces and undergo similar cycles of proliferation and shedding as normal endometrium. It is hypothesised that these fragments may contain endometrial stem cells that are abnormally shed during menses and have the ability to implant into ectopic surfaces and proliferate forming endometriotic lesions [16, 17]. Although endometrial stem cells have been identified in menstrual blood, they have not been recognised in peritoneal fluid in higher quantities during menstruation [18]. This may be due to the lack of structural markers for these cells.

Infrared (IR) spectroscopy is a powerful technique to investigate biological tissues, since it can detect many important biochemical signatures including Amide I (~1,650 cm^{-1}), Amide II (~1,550 cm^{-1}), protein (~1,425 cm^{-1}), Amide III (~1,260 cm^{-1}), asymmetric phosphate stretching vibrations ($\nu_{as}PO_2^-$; ~1,225 cm^{-1}), carbohydrates (~1,155 cm^{-1}), symmetric phosphate stretching vibrations ($\nu_sPO_2^-$; ~1,080 cm^{-1}) and protein phosphorylation (~970 cm^{-1}) [19]. Successful applications of IR spectroscopy towards the analysis of tissues samples includes cancer identification in breast, lung, colon and prostate tissues [20, 21]. Endometrial tissues have been also investigated using IR spectroscopy, where results for differentiating benign and malignant tissues were mainly assigned to lipid and Amide I / II regions [22]. Putative stem cells have been studied in human intestinal crypts using IR spectroscopy where $\nu_sPO_2^-$ were found to be the main biomarker for distinguishing different putative cell types [23].

Among different types of IR techniques, synchrotron radiation-based Fourier-transform infrared (SR-FTIR) and focal plane array (FPA) FTIR spectroscopy are excellent techniques for investigating tissues samples [21]. In SR-FTIR, a synchrotron source emits a collimated light beam more intense than that of a bench-top

spectrometer. This provides an excellent signal-to-noise ratio (SNR) that is 1000-times greater to that of conventional IR sources and allows spatial resolutions as small as 10 μm [24, 25]. The absence of thermal noise and the order of intensity magnitude greater for synchrotron radiation source in infrared microspectroscopy increased the SNR performance in comparison to operation with a built-in globar (thermal) source; therefore, generating a spectra with higher resolution than regular FTIR [21]. On the other hand, FPA uses an IR focal plane array detector to generate hyperspectral imaging. The multiple detector elements of the FPA detector enable the concurrent acquisition of several spectra at each spatial point of the area of the sample under investigation. This allows the examination of larger areas at reasonable experimental time periods. This method results in the formation of a "hypercube" which contains information in two spatial dimensions: a pseudo-image and one spectral dimension, corresponding to the spectrum for each point (pixel) of that image [21].

Biospectroscopy is a powerful technology that can be used in conjunction to other biomarker-based isolation techniques to attempt the identification of endometrial stem cells amongst their progeny. The characterisation of endometrial stem cells with regards to their molecular, genetic and epigenetic make-up may assist in unravelling their role in many endometrial proliferative disorders such as endometriosis, adenomyosis and carcinogenesis. In this paper, SR-FTIR and FPA-FTIR microispectroscopy were used to investigate endometrial tissues samples. SR-FTIR microspectroscopy was used for identification of putative stem cells among the epithelial cells on the basalis portion of endometrial glands; while, FPA-FTIR microspectroscopy was used to distinguish epithelial cells lining in the basalis portion from those lining in the functionalis portion of the endometrial glands.

Methods

Tissue collection and preparation

Ethical approval was obtained from Liverpool Adult Research Ethics Committee (LREC 09/H1005/55 and 11/H1005/4), and the study was conducted according to the principles of the Declaration of Helsinki and all other applicable national or local laws and regulations. Written, informed consent was obtained from all participants prior to inclusion in the study at the Liverpool Women's NHS Foundation Trust.

Three secretory phase pre-menopausal uteruses from women undergoing hysterectomy for non-endometrial benign causes (uterine prolapse) who had not been on hormonal treatments in the preceding 3 months were collected. A wedge of tissue from the lumen to the muscular myometrial layer that included superficial and basal endometrium as well as myometrium was taken from the detached uterus, collected in normal buffered formalin and subsequently embedded in paraffin. Several 10-µm-thick parallel tissue sections were cut from each sample and floated onto 1 cm × 1 cm BaF_2 slides (Photox Optical Systems). These were de-waxed by serial immersion in 3 sequential fresh xylene baths for 5 minutes and washed in an acetone bath for a further 5 minutes. The resulting samples were allowed to air dry and placed in a desiccator until analysis. Four-µm-thick parallel tissue sections were floated to glass slides and stained with haematoxylin and eosin (H&E) for histological comparison.

Synchrotron radiation-based Fourier-transform infrared (SR-FTIR) microspectroscopy

The SR-FTIR microspectroscopy setup is illustrated in Fig. 1. Spectral mapping data from 4,000 cm^{-1} to 600 cm^{-1} were obtained at the Diamond light Source Ltd, UK

(www.diamond.ac.uk) at beamline 22IR. Optically interfaced to the beamline were a Bruker Vertex 80v FTIR spectrometer, a Bruker Hyperion 3000 microscope equipped with front surface reflecting optics, a liquid nitrogen mercury cadmium telluride (MCT) detector and a 36× objective lens. Spectra were collected in transmission mode from specimens onto BaF_2 slides. At 10-µm step sizes, 256 spectra were co-added with the aperture set for 10 µm × 10 µm (numerical aperture [NA] = 0.50, lateral resolution 6.7–12.2 µm). Mapping data were acquired within 6 h (Fig. 1a) with background spectra updated after every 10 spectra. Spectra recorded in transmittance were converted and displayed as absorbance with the internal Bruker Optics OPUS 8 software (Bruker Optics, Ettlingen, Germany).

[Insert Figure 1 here]

Thermal source focal plane array Fourier-transform infrared (FPA-FTIR) microspectroscopy

A Bruker Vertex 80v FTIR spectrometer and Hyperion 3000 microscope were employed using the built-in globar (thermal) infrared source. A 64 × 64 cooled photodiode FPA collected spectra from 4,000-600 cm^{-1} in parallel at 4 cm^{-1} resolution (NA = 0.65, lateral resolution 5.2–9.4 µm). Co-addition of 128 sample scans and 256 background scans was used. Matching 15× Cassegrain objective and condenser mirror lenses were used for transmission measurement of microtomed tissue sections mounted on BaF_2 slides (1 cm × 1 cm). FPA imaging maps were acquired from several sections from the three uterine samples involving basalis/functionalis junctions containing epithelial glands. The fidelity of the FPA spectral images region between 1,800 cm^{-1} and 900 cm^{-1} at 4 cm^{-1} spectral resolution (468 wavenumbers with data

spacing of 1.9 cm^{-1}) is the area associated with the biological spectral fingerprint [26] (Figure 2).

Data pre-processing

Absorbance spectral images were converted to suitable digital files (.txt) for analysis within MATLAB R2014a environment (Mathworks Inc., Natick, USA). They included wavenumber regions scanned between 4000 cm^{-1} and 600 cm^{-1} that were truncated to include only the fingerprint region between 1,800 and 900 cm^{-1}. The resulting dataset was smoothed by Savitzky-Golay filter, rubber band baseline-corrected and normalized to the Amide I peak (*i.e.*, ≈1,650 cm^{-1}) [21, 27]. The normalization to Amide I peak is commonly performed after baseline correction in order to correct spectra having different absorbance values at Amide I region [21]. This prevents the effect of samples having different thickness or concentration that can mask the actual biochemical difference of interest [27]. The importing and pre-treatment of the spectral data and the construction of chemometric classification models were performed using PLS Toolbox 7.8 (Eigenvector Research Inc., Wenatchee, USA) within MATLAB software.

Computational analysis

Computational analysis consisted of three models: principal component analysis (PCA), successive projections algorithm (SPA) and genetic algorithm (GA). All models were followed by linear discriminant analysis (LDA) [28]. Before applying each analytical model, spectral data were divided into training (70%), validation (15%) and prediction (15%) sets by applying the classic Kennard-Stone (KS) uniform sampling algorithm [29]. The training set was dependent on the number of pixels of the FPA detector for the same IR source; whereas the prediction set was only used for the final classification evaluation.

PCA is a multivariate analysis technique that aims to reduce the number of variables present in the spectral dataset [30]. Principal components (PCs) capture most of the variance present in the original dataset. After PCA decomposition, the original data is represented by a combination of scores and loadings. The scores represent the variance on the samples direction, therefore they show the similarities and dissimilarities among the samples; and the loadings represent the variance on the variables direction, therefore they show the weight of each wavenumber for the pattern observed on the scores. The scores are used as input for LDA discriminant analysis in PCA-LDA [31]. In addition, after PCA it is possible to obtain the Q residuals value calculated as the sum of the squares of the residual values at each variable for each sample (squared prediction error). This indicates the similarity of the residuals between the samples and its projection into the principal component space, representing the lack of fit in the PCA model. This is therefore a tool for outlier detection [32].

SPA is a forward feature selection method [33], which operates solving co-linearity problems by selecting wavelengths whose information content is minimally redundant. The model is built through a series of interactions, starting with one wavelength and then incorporating a new one at each iteration until a specified number (N) of wavelengths is reached. Differently from PCA, SPA does not modify the original data space, since the projections are used only for selection purposes; thus, the relationship between spectral variables and the original data space is preserved [34].

Genetic algorithms are combinational algorithms inspired by Mendelian genetics. They use a combination of selection, recombination and mutation to evolve a solution to a problem. They treat data as chromosomes allocating reproductive

opportunities in such a way that those chromosomes, which represent a better solution to the target problem are given more chances to "reproduce" than those, which represent poorer solutions [34, 35]. The GA routine was carried out using 100 generations containing 200 chromosomes each. Crossover and mutation probabilities were set to 60% and 10%, respectively. Moreover, the algorithm was repeated three times, starting from different random initial populations. The best solution, in terms of the fitness value, resulting from the three realizations of the GA was employed.

LDA was performed following the application of each of the analytical models. LDA scores, loadings, and discriminant function (DF) values were obtained. Usually, the first LDA factor is used to visualize the main biochemical alterations within the sample on a 1-dimensional scores plot. The optimum number of variables for SPA-LDA and GA-LDA was determined by the minimum cost function G calculated for a given validation dataset as:

$$G = \frac{1}{N_V}\sum_{n=1}^{N_V} g_n \quad (1)$$

where N_V is the number of validation samples and g_n is defined as:

$$g_n = \frac{r^2(x_n, m_{I(n)})}{\min_{I(m)\neq I(n)} r^2(x_n, m_{I(m)})} \quad (2)$$

in which $r^2(x_n, m_{I(n)})$ is the squared Mahalanobis distance between object x_n (of class index $I(n)$) and the centre of its true class ($m_{I(n)}$); and $r^2(x_n, m_{I(m)})$ is the squared Mahalanobis distance between object x_n and the centre of the closest wrong class ($m_{I(m)}$) [36].

For each model, sensitivity (the confidence in a positive result for a sample of the label class is obtained) and specificity (the confidence that a negative result for a sample of non-label class is obtained) were calculated as important quality standards

in test evaluation. The quality metrics used in this study for evaluating the classification results can be calculated following the equations [36]:

$$\text{Sensitivity}(\%) = \frac{\text{TP}}{\text{TP+FN}} \times 100 \quad (3)$$

$$\text{Specificity}(\%) = \frac{\text{TN}}{\text{TN+FP}} \times 100 \quad (4)$$

where FN is defined as false negative, FP as false positive, TP as true positive and TN as true negative.

Results

Differences between basalis and functionalis for all specimens by FPA imaging

In order to investigate the presence of spectral differences between the functionalis and basalis epithelial layers all specimens were interrogated using FPA imaging. A raster scan approach was applied to include the epithelial layers surrounding the glandular base within the specimens. Micrographs of the involved areas with overlaid markers were used for identification of the specific epithelial regions from which spectra were acquired (Fig. 2b and 2d). Visual correlation with parallel H&E tissue sections was used for selection of different epithelial regions (Fig. 2a and 2c). Chemometric analyses of the location-derived spectra (Fig. 2e) allowed classification into basalis and functionalis regions.

[Insert Figure 2 here]

The chemometric analyses employed for the differentiation of the epithelial cells residing in the two regions included PCA-LDA, SPA-LDA and GA-LDA that were applied to the pre-processed dataset. In total, n = 65,832 spectra resided within the functionalis class, and n = 38,357 spectra within the basalis class for the 18 areas were interrogated for comparison. Fig. 2f shows the pre-processed spectra with the

basalis area depicted in blue and the functionalis area in red. Overall, the IR spectra appear to have similarities in the biochemical fingerprint region (1,800 cm^{-1} to 900 cm^{-1}). Fig. 3a shows the explained variance for each principal component (PC) using PCA-LDA. Eight PCs (explained variance = 89.41%) were used for model construction. Fig. 3b shows the 1-D scores plot derived by PCA-LDA. It reveals some segregation of the basalis from the functionalis, although with some overlap of scores along DF1. The majority of the difference between the two regions was attributed to absorptions at 1,540 cm^{-1} and 1,625 - 1,675 cm^{-1} according to the PCA loadings [see Electronic Supplementary Information (ESI) Fig. S1a].

[Insert Figure 3 here]

Although GA-LDA classified very successfully the two classes using 14 variables determined from the minimum cost function G (Fig. 3e) (see ESI Table S2), in which the related 1-D scores plot from different classes dissociated along DF1 is shown in Fig. 3f, the chemometric technique that classified the 2 classes most successfully was SPA-LDA using 12 variables (Fig. 3c) (see ESI Table S2), since better classifying performance was achieved using this technique in the prediction set (Table 1). Also, the related 1-D scores plot illustrates that spectral points from different classes dissociate very well along DF1 (Fig. 3d).

[Insert Table 1 here]

All three techniques identified differences that aided classification within similar spectral regions. These differences were tentatively identified in the spectral regions of 1,650 cm^{-1} (Amide I band) and 1510 cm^{-1} (C-N stretch contribution to the Amide II vibrational band). The overall classification rates by each chemometric

technique employed to differentiate the basalis from the functionalis using FPA spectral data is shown in ESI Table S3.

Differences between basalis and functionalis in different individuals by FPA imaging

To investigate whether similar spectral areas are able to identify segregation between the basalis and functionalis regions in different individuals, the same chemometric methods were applied on each of the specimens separately. We used the same fingerprint spectral regions for each of the specimens (see ESI Fig. 2a, 3a and 4a). When the average of the spectra derived from the basalis layer were compared with those from the functionalis layers, subtle differences were visualised (see ESI Fig. S2b, S3b and S4b). The classification rates produced by the different chemometric techniques (see ESI Table S4) varied from 50.0 to 100.0% in the prediction set. The wavenumbers used for SPA-LDA and GA-LDA in this case are presented in ESI Table S5 and graphically in ESI Fig. S2g and S2e, S3g and S3e and S4g and S4e, respectively. PCA loadings for these specimens are shown in ESI Fig. S1b, c and d, in which it is observed that most of the differences between the two regions are associated with absorptions on 1,235, 1,550 and 1,628 - 1,685 cm^{-1}. Associated scores plots graphically represent the separation identified between the two layers utilising the three chemometric techniques (see ESI Fig. S2, 3, 4: d, f and h). Similar spectral areas seem to be causing segregation of the two histologic regions when specimens are analysed separately. These areas mostly include the spectral regions of 1,650 cm^{-1} (Amide I band) and 1,510 cm^{-1} (C-N stretch contribution to the Amide II vibrational band). These spectral areas correspond to those responsible for class segregation when all specimens were examined together. This mean that spectral

variations causing segregation of the two histological areas are constant independent of other inter-individual variability that may exists within the sample pool.

Inter-individual differences within the basalis layer by FPA imaging

To investigate if the basalis layer is dissimilar in different individuals PCA-LDA, SPA-LDA and GA-LDA approaches were applied to the pre-processed spectra derived from the basalis regions of all specimens (see ESI Fig. S5a). ESI Fig. S5b shows the average spectrum taken from the basalis layer of each individual. The 2-D scores plots in ESI Fig. S5d, f and h, all show clustering of spectra from each individual as well as segregation of the spectra from different specimens. PCA-LDA identified 8 PCs (explained variance = 92.61%) as the main responsible for this variation (ESI Fig. S5c). SPA-LDA utilised 12 wavenumbers (ESI Fig. S5e) and GA-LDA utilised 11 wavenumbers (ESI Fig. S5g). The classification rates between the specimens within basalis layer are shown on ESI Table S6. ESI Table S7 shows the wavenumbers used for the analyses using SPA-LDA and GA-LDA. Wavenumbers between 1,600 cm^{-1} and 1,700 cm^{-1} are consistently responsible for the segregation between the basalis layers of different specimens as well as for the classification into basalis and functionalis layers as show above.

Investigation of epithelial variability at the base of the glands using synchrotron radiation

The spatial resolution of synchrotron-based FTIR microspectroscopy is in the region of 10 μm × 10 μm. The high resolution allowed the investigation of differences that may exist between epithelial cells that reside close to the base of the glands within the basalis layer. A micrograph of the involved areas with overlaid markers was used for identification of the specific regions of which spectra were acquired. Spectra were

acquired along glands, which were cut longitudinally ensuring that the deepest portion of the gland was examined (Fig. 4a). The derived spectra were divided into 3 classes according to their position (Fig. 4b). Class 1 included spectra taken closer to the deeper margins of the glands with the other classes being sequentially more superficial. Class 1 contained 60 spectra, Class 2 contained 310 spectra and Class 3 contained 30 spectra. PCA-LDA, SPA-LDA and GA-LDA of the three classes were employed using 2 wavenumbers for SPA and 20 for GA (Fig. 4c, e and g). Initial attempts to classification of these classes were underpowered as revealed by the 2-D scores plots in Fig. 4d, f and h, where segregation of classes is not readily visible, despite noticeable progression of the scores especially in the GA-LDA-derived plot.

[Insert Figure 4 here]

To identify whether there is any variation between the epithelial cells resting at the terminal end of the glands, the same analysis was employed comparing the spectra derived from the classes from above that were furthest apart (Classes 1 and 3) (Fig. 5a and b). Interestingly, there was separation between these two classes when compared by all three chemometric techniques as evidenced by the 2-D scores plots in Fig. 5d, f and h, especially using GA-LDA. In this case, 8 PCs (explained variance = 93.72%) were used for PCA-LDA, 2 wavenumbers were utilised for SPA-LDA and 8 for GA-LDA (Fig. 5c, e, g and ESI Table S8), being predominantly close to the region of ~1,234 cm^{-1}, ~1,550 cm^{-1} and ~1,650 cm^{-1} which corresponds to the biomarkers responses from Amide III, Amide II and Amide I, respectively [26]. The classification rates for each technique in this case are presented in ESI Table S9.

[Insert Figure 5 here]

Synchrotron radiation-based IR spectra for the identification of the location of stem cells

To investigate the existence of putative stem cells within the deepest/terminal portion of the glands, we searched for the existence of spectral discriminating factors between all epithelial cells in this area. Each spectrum was treated individually but spectra were also placed into classes (1 to 8) depending on their distance from the deepest part of the gland (Fig. 6a). Ten adjacent spectra were sequentially placed in each class. The reason for this was to attempt the localisation of putative stem cells in addition to their existence. The original spectra (Fig. 6b) were cut to only include the biological fingerprint region and were pre-processed as before (Fig. 6c). They underwent exploratory PCA using the first 3 PCs (explained variance = 99.61%). Two-dimensional scores plots were extracted representing the 2 first PCs (Fig. 6d). Each point on the resulting scores plot represented a single point on the image maps. Using a 95% confidence interval (CI), spectral points that segregated from the clustered spectra were identified. These "outliers" were particularly obvious on the PC2 axis and included mostly cells from Classes 1, 2 and 5 and 6. Interestingly, when Q residuals analysis was applied, Classes 1 and 5 also featured outlier spectra (Fig. 6e). Following analysis by these unsupervised techniques, spectra were classified depending on whether they were identified as outliers or not (Fig. 6f). The averaged spectra resulting from these two classes ("normal" and "outliers") exposed visible differences (Fig. 6g). A resulting subtraction spectrum identified differences in the absorption bands near 1,080 cm^{-1} (symmetric PO_2^- stretching vibrations in RNA and DNA), 1,390 cm^{-1} (nucleic acids), 1,550 cm^{-1} (Amide II) and 1,650 cm^{-1} (Amide I) (Fig. 6h).

[Insert Figure 6 here]

These absorption bands were utilized to produce intensity images based on the FPA spectra to show where in the gland their intensity is maximal to attempt visualization of "outlier" cellular regions that include putative stem cells (Fig. 7). Although the definition of the FPA images does not allow precise localization of the highest intensities it is broadly evident that there are in general three areas within the gland where these outlier cells reside: these are along the deepest portion of the gland (Fig. 7a – circle in blue), the basalis/functionalis junction (Fig. 7a – circled in red) and close to the most superficial part of the gland opening into the endometrial cavity (Fig. 7a – circled in black). The stem cells were found within the population of epithelial cells lining the basalis portion of endometrial glands.

[Insert Figure 7 here]

Discussion

This study employed powerful chemometric techniques in conjunction with synchrotron radiation and globar IR technologies to classify epithelial histological areas within the endometrial glands. It also explored the potential of such technologies in isolating putative stem cells and determining their location of within the endometrial epithelium. Since the progress in characterising the human endometrial epithelial stem cells has been particularly poor, our method may attract the attention of endometrial biologists.

Pathophysiological morphological and molecular alterations of epithelial cells that line the glands in the basalis and functionalis layers may be associated with both benign and malignant diseases arising from the endometrium. Therefore,

understanding the differences in molecular composition of the cells within the examined histological areas may assist directing research to further pinpoint the exact locations of such pathological changes and to potentially formulate curative strategies.

Computational analysis of the obtained spectra identified significant differences between the epithelial cells lining the glands in the basalis and functionalis regions. Some of the discriminating wavenumbers for this variation remained constant when uterine specimens from different individuals were examined separately. Namely Amide I and stretching of C=N bonds associated spectral bands correctly classified the two histological regions. This molecular variability between the two layers may assist with the further characterisation of the basalis epithelial cells, which are postulated to give rise to many endometrial proliferative conditions such as endometriosis. The functionalis layer is shed during the menstrual shedding leaving the basalis from which the new functionalis is thought to be regenerated. Noe *et al.* [3] have suggested that the cells from the basalis layer are responsible for the establishment of ectopic endometriotic lesions, and that is further supported by the baboon model of induction of endometriosis [17]. In the baboons, the endometrium biopsied during day 2 of the menses (when the basalis is easily accessible) when instilled in the pelvic cavity give rise to identical endometriotic deposits to humans. This is likely to be due to the enrichment of stem cells in the basalis layer, that is sampled during menses and these cells are likely to survive and establish endometriotic lesions. Therefore, identifying such discriminatory spectroscopic biomarkers can also be significant due to their potential for translation into clinical practice. Biomarkers identified using the above techniques may be further explored to interrogate the basalis endometrial layer of women with and without endometriosis.

Several studies have attempted the identification of adult stem cells within the basalis layer [1, 6]. These studies have utilised different stem cell specific activities for identification [9, 10, 13]. Our study intended to isolate and locate putative epithelial stem cells using morphological, molecular and chemical variability that exists within the cells lining the glandular glands. There is expanding evidence that FTIR microspectroscopy is capable of identifying stem cells in several tissues including cornea, epidermis and intestine [23, 37-40]. The chemometric analysis of the IR data we acquired was able to identify specific putative stem cell locations within the population of epithelial cells. Both FPA and synchrotron radiation -based techniques identified putative stem cells at the deepest part of the gland and at a distance of about 50 μm from that location. This indicates that there are at least two sets of stem cells residing in these gland bases. We speculate that the stem cells have different functions: one being dormant and the other actively differentiating to more specialised daughter cells.

Analysis of the synchrotron radiation data has also shown mild progressive differences between the epithelial cells lining the bases of the glands. Although when the synchrotron radiation-derived spectral data from these cells were compared by dividing them into three classes their variability was minimal, when comparing the classes positioned furthest apart they exhibited significant differences. This points to a progressive alteration in the molecular structure of epithelial cells even in the basalis layer.

All samples in our study were obtained from women that had hysterectomies for reasons other than endometrial pathology or malignancy. Our aim was to perform preliminary spectroscopic investigations on these tissues to evaluate the performance of these techniques

in identifying different cellular linages within the endometrial glands. The variability between endometrial samples depends on cycle phase, hormonal factors, lifestyle, body mass index, weight, diet and alcohol consumption, parity and several other factors, therefore obtaining a completely homogeneous endometrial samples from different individuals is virtually impossible. Although our sample size was small, all 3 women included were sampled within the secretory phase of the cycle to remove cycle dependent variability and all women had histologically confirmed healthy endometrium without any prior hormonal treatments. The three different samples in our study were used to support information derived when using all specimens together.

Identification of endometrial adult stem cells may be of value in all endometrial proliferative disease including carcinogenesis research as they may be involved in related processes by their alteration into cancer stem cells [15]. Their localisation by means of specific biomarker as extracted by spectroscopic techniques may in the future provide diagnostic and therapeutic avenues for uterine cancer.

Conclusion

With this study, we demonstrate that SR-FTIR microspectroscopy coupled with multivariate computational analysis and wavenumber selection techniques might be able to identify putative stem cells within the population of epithelial cells lining the basalis portion of endometrial glands. Specific identification of the location of these cells remains elusive but potential stem cells were segregated from their surrounding cells based on their inherent IR spectroscopic signatures. Specific spectral biomarkers responsible for this segregation were identified and corresponded mainly to PO_2^- vibrational modes of DNA and RNA, nucleic acids and Amide I and II. In addition, FPA-FTIR spectroscopy coupled with similar chemometric analysis was able to consistently differentiate the epithelial cells lining the

basalis portion of the endometrial glands from those lining the functionalis portion. Further future developments in both spectroscopic technologies and related chemometric techniques may be able to track cellular lineages and individual cells within all tissues therefore potentially revealing physiological and pathological tissue functions based on morphological alterations.

Acknowledgements

The authors would like to thank Rosemere Cancer Foundation for financial support; and the Lancashire Teaching Hospitals, Liverpool Women's NHS Foundation Trusts and the Diamond Light Source Research Facility for all support related to this study. CLMM would like to thank CAPES-Brazil (Doutorado Pleno no Exterior, grant 88881.128982/2016-01) for financial support; and KMGL thanks CNPq (grant 305962/2014-0) for financial support.

Conflict of Interest: The authors declare no competing financial interests and no other conflict of interest.

List of Abbreviations

DF: discriminant function;

DNA: deoxyribonucleic acid;

FN: false negative;

FP: false positive;

FPA: focal plane array;

FPA-FTIR: focal plane array-based Fourier-transform infrared;

FTIR: Fourier-transform infrared;

GA: genetic algorithm;

GA-LDA: genetic algorithm – linear discriminant analysis;

H&E: haematoxylin and eosin;

IR: infrared;

KS: Kennard-Stone;

LDA: linear discriminant analysis;

MCT: mercury cadmium telluride;

NA: numerical aperture;

PCA: principal component analysis;

PCA-LDA: principal component analysis – linear discriminant analysis;

PCs: principal components;

RNA: ribonucleic acid;

SNR: signal-to-noise ratio;

SPA: successive projections algorithm;

SPA-LDA: successive projections algorithm – linear discriminant analysis;

SR-FTIR: synchrotron radiation-based Fourier-transform infrared;

TN: true negative;

TP: true positive.

References

1. Gargett CE, Chan RW, Schwab KE. Endometrial stem cells. Curr Opin Obstet Gynecol. 2007;19:377-83.
2. Hapangama DK, Kamal AM, Bulmer JN. Estrogen receptor β: the guardian of the endometrium. Hum Reprod Update. 2015;21:174-93.
3. Noe M, Kunz G, Herbertz M, Mall G, Leyendecker G. The cyclic pattern of the immunocytochemical expression of oestrogen and progesterone receptors in human myometrial and endometrial layers: characterization of the endometrial-subendometrial unit. Hum Reprod. 1999;14:190-7.
4. Li L, Xie T. Stem cell niche: structure and function. Annu Rev Cell Dev Biol. 2005;21:605-31.
5. Eckfeldt CE, Mendenhall EM, Verfaillie CM. The molecular repertoire of the 'almighty' stem cell. Nat Rev Mol Cell Biol. 2005;6:726-37.
6. Gargett CE, Masuda H. Adult stem cells in the endometrium. Mol Hum Reprod. 2010;16:818-34.
7. Patel AN, Park E, Kuzman M, Benetti F, Silva FJ, Allickson JG. Multipotent menstrual blood stromal stem cells: isolation, characterization, and differentiation. Cell Transplant. 2008;17:303-11.
8. Musina RA, Belyavski AV, Tarusova OV, Solovyova EV, Sukhikh GT. Endometrial mesenchymal stem cells isolated from the menstrual blood. Bull Exp Biol Med. 2008;145:539-43.
9. Chan RW, Schwab KE, Gargett CE. Clonogenicity of human endometrial epithelial and stromal cells. Biol Reprod. 2004;70:1738-50.

10. Gargett CE, Schwab KE, Zillwood RM, Nguyen HP, Wu D. Isolation and culture of epithelial progenitors and mesenchymal stem cells from human endometrium. Biol Reprod. 2009;80:1136-45.
11. Kim JY, Tavare S, Shibata D. Counting human somatic cell replications: methylation mirrors endometrial stem cell divisions. Proc Natl Acad Sci USA. 2005;102:17739-44.
12. Valentijn AJ, Palial K, Al-lamee H, Tempest N, Drury J, Von Zglinicki T, Saretzki G, Murray P, Gargett CE, Hapangama DK. SSEA-1 isolates human endometrial basal glandular epithelial cells: phenotypic and functional characterization and implications in the pathogenesis of endometriosis. Hum Reprod. 2013;28:2695-708.
13. Masuda H, Kalka C, Takahashi T, Yoshida M, Wada M, Kobori M, Itoh R, Iwaguro H, Eguchi M, Iwami Y, Tanaka R, Nakagawa Y, Sugimoto A, Ninomiya S, Hayashi S, Kato S, Asahara T. Estrogen-mediated endothelial progenitor cell biology and kinetics for physiological postnatal vasculogenesis. Circ Res. 2007;101:598-606.
14. Götte M, Wolf M, Staebler A, Buchweitz O, Kelsch R, Schüring AN, Kiesel L. Increased expression of the adult stem cell marker Musashi-1 in endometriosis and endometrial carcinoma. J Pathol. 2008;215:317-29.
15. Rosen JM, Jordan CT. The increasing complexity of the cancer stem cell paradigm. Science. 2009;324:1670-73.
16. Sasson IE, Taylor HS. Stem cells and the pathogenesis of endometriosis. Ann N Y Acad Sci. 2008;1127:106-15.
17. Sourial S, Tempest N, Hapangama DK. Theories on the pathogenesis of endometriosis. Int J Reprod Med. 2014;2014:179515.
18. Bokor A, Debrock S, Drijkoningen M, Goossens W, Fülöp V, D'Hooghe T. Quantity and quality of retrograde menstruation: a case control study. Reprod Biol Endocrinol. 2009; 7:123.

19. Siqueira LFS, Lima KMG. A decade (2004 – 2014) of FTIR prostate cancer spectroscopy studies: An overview of recent advancements. Trends Analyt Chem. 2016;82:208-21.

20. Siqueira LFS, Lima KMG. MIR-biospectroscopy coupled with chemometrics in cancer studies. Analyst. 2016;141:4833-47.

21. Baker MJ, Trevisan J, Bassan P, Bhargava R, Butler HJ, Dorling KM, Fielden PR, Fogarty SW, Fullwood NJ, Heys KA, Hughes C, Lasch P, Martin-Hirsch PL, Obinaju B, Sockalingum GD, Sulé-Suso J, Strong RJ, Walsh MJ, Wood BR, Gardner P, Martin FL. Using Fourier transform IR spectroscopy to analyze biological materials. Nat Protoc. 2014;9:1771-91.

22. Taylor SE, Cheung KT, Patel II, Trevisan J, Stringfellow HF, Ashton KM, Wood NJ, Keating PJ, Martin-Hirsch PL, Martin FL. Infrared spectroscopy with multivariate analysis to interrogate endometrial tissue: a novel and objective diagnostic approach. Br J Cancer. 2011;104:790-97.

23. Walsh MJ, Fellous TG, Hammiche A, Lin W-R, Fullwood NJ, Grude O, Bahrami F, Nicholson JM, Cotte M, Susini J, Pollock HM, Brittan M, Martin-Hirsch PL, Alison MR, Martin FL. Fourier transform infrared microspectroscopy identifies symmetric PO_2^- modifications as a marker of the putative stem cell region of human intestinal crypts. Stem Cells. 2008;26:108-18.

24. Tobin MJ, Chesters MA, Chalmers JM, Rutten FJ, Fisher SE, Symonds IM, Hitchcock A, Allibone R, Dias-Gunasekara S. Infrared microscopy of epithelial cancer cells in whole tissues and in tissue culture, using synchrotron radiation. Faraday Discuss. 2004;126:27-39.

25. Miller LM, Dumas P. Chemical imaging of biological tissue with synchrotron infrared light. Biochim Biophys Acta. 2006;1758:846-57.

26. Movasaghi Z, Rehman, S, ur Rehman I. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. Applied Spectroscopy Reviews. 2008;43:134-79.

27. Trevisan J, Angelov PP, Carmichael PL, Scott AD, Martin FL. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. Analyst 2012;137:3202-15.

28. Martin FL, Kelly JG, Llabjani V, Martin-Hirsch PL, Patel II, Trevisan J, Fullwood NJ, Walsh MJ. Distinguishing cell types or populations based on the computational analysis of their infrared spectra. Nat. Protoc. 2010;5:1748-60.

29. Kennard RW, Stone LA. Computer adided design of experiments. Technometrics. 1969;11:137-48.

30. Bro R, Smilde AK. Principal component analysis. Anal Methods. 2014;6:2812-31.

31. Morais CLM, Lima KMG. Principal component analysis with linear and quadratic districiminant analysis for identification of cancer samples based on mass spectrometry. J Braz Chem Soc. 2018;29:472-81.

32. Keithley RB, Heien ML, Wightman RM. Multivariate concentration determination using principal component regression with residual analysis. Trends Analyt Chem. 2009;28:1127-36.

33. Soares SFC, Gomes AA, Araujo MCU, Filho ARG, Galvão RKH. The successive projections algorithm. Trends Analyt Chem. 2013;42:84-98.

34. Theophilou G, Lima KMG, Briggs M, Martin-Hirsch PL, Stringfellow HF, Martin FL. A biospectroscopic analysis of human prostate tissue obtained from different time periods points to a trans-generational alteration in spectral phenotype. Sci. Rep. 2015;5:13465.

35. Whitley D. A Genetic Algorithm Tutorial. Satistics and Computing. 1994;4:65-85.

36. Siqueira LFS, Araújo Júnior RF, Araújo AA, Morais CLM, Lima KMG. LDA vs. QDA for FT-MIR prostate cancer tissue classification. Chemometr Intell Lab Syst. 2017;162:123-9.

37. Kelly JG, Nakamura T, Kinoshita S, Fullwood NJ, Martin FL. Evidence for a stem-cell lineage in corneal squamous cell carcinoma using synchrotron-based Fourier-transform infrared microspectroscopy and multivariate analysis. Analyst 2010;135:3120-25.

38. Fogarty SW, Patel II, Trevisan J, Nakamura T, Hirschmugl CJ, Fullwood NJ, Martin FL. Sub-cellular spectrochemical imaging of isolated human corneal cells employing synchrotron radiation-based Fourier-transform infrared microspectroscopy. Analyst 2013;138:240-48.

39. Patel II, Harrison WJ, Kerns JG, Filik J, Wehbe K, Carmichael PL, Scott AD, Philpott MP, Frogley MD, Cinque G, Martin FL. Isolating stem cells in the inter-follicular epidermis employing synchrotron radiation-based Fourier-transform infrared microspectroscopy and focal plane array imaging. Anal. Bioanal. Chem. 2012;404:1745-58.

40. Walsh MJ, Hammiche A, Fellous TG, Nicholson JM, Cotte M, Susini J, Fullwood NJ, Martin-Hirsch PL, Alison MR, Martin FL. Tracking the cell hierarchy in the human intestine using biochemical signatures derived by mid-infrared microspectroscopy. Stem Cell Res. 2009;3:15-27.

Table 1. Predictive performance presented as sensibility and sensitivity rates calculated for each chemometric model developed to differentiate basalis and functionalis tissues from the FPA-FTIR spectral data.

	PCA-LDA	SPA-LDA	GA-LDA
Basalis			
Sensitivity (%)	66.7	100.0	75.0
Specificity (%)	66.7	100.0	100.0
Functionalis			
Sensitivity (%)	0.0	100.0	60.0
Specificity (%)	0.0	100.0	100.0

Figure captions

Figure 1. Synchrotron radiation-based FTIR spectral interrogation and analysis of endometrial epithelial gland bases to identify the location of putative stem cells. **(a)** Micrograph with superimposed marks representing areas interrogated by synchrotron radiation (magnification 36×); **(b)** Diagram representing a Michelson interferometer, which is the basic structure of an IR spectroscope; **(c)** The interferogram is transformed to an absorbance spectrum by Fourier-transform; and, **(d)** PCA identifies "outliers" which may represent putative stem cells.

Figure 2. Representation of region identification for spectral selection by FPA and pre-processing of acquired spectra. **(a)** H&E-stained parallel section of examined longitudinally cut crypt, in the secretory phase (circled in blue: cells in the deepest portion of the gland; circled in red: basalis/functionalis cells; circled in black: cells close to the gland opening into the endometrial cavity); **(b)** Micrograph of examined longitudinally cut crypt with overlaid circular markers representing epithelial areas which were selected for comparison (magnification 15×); (the actual spectral area selection on the pseudo-image was performed at different magnifications to ensure the spectra selected were good representatives of the areas to be examined); **(c)** H&E-stained parallel section of examined coronally cut crypt, in the secretory phase; **(d)** Micrograph of examined coronally cut crypt with overlaid circular markers representing epithelial areas which were selected for comparison (magnification 15×) (the actual spectral area selection on the pseudo-image was performed at different magnifications to ensure the spectra selected were good representatives of the areas to be examined); **(e)** Averaged absorbance spectra acquired by FPA imaging of all samples; and, **(f)** Pre-processed spectra resulting from the raw spectra acquired by FPA imaging of all samples.

Figure 3. Classification of Basalis and Functionalis regions by spectral analysis using PCA-LDA, SPA-LDA and GA-LDA on FPA-FTIR derived data (red = Functionalis, blue = Basalis): All specimens. **(a)** Cost/function plot identifying the optimal number of PCs to be used for PCA; **(b)** Scores plot graphically representing classification by PCA-LDA. The *x*-axis represents the sample index and the *y*-axis the discriminant factor 1 (DF1); **(c)** Wavenumber selection for SPA-LDA; **(d)** Scores plot graphically representing classification by SPA-LDA. The *x*-axis represents the sample index and the *y*-axis DF1; **(e)** Wavenumber selection for GA-LDA; **(f)** Scores plot graphically representing classification by GA-LDA. The *x*-axis represents sample index and the *y*-axis DF1.

Figure 4. Classification of epithelial cells according to their position in the endometrial crypt base by synchrotron radiation spectral analysis using PCA-LDA, SPA-LDA and GA-LDA on FPA-FTIR microscopy-derived spectral data. **(a)** Micrograph with circles representing the regions sampled (this is a representation as the areas analysed were selected at different magnifications to reveal the synchrotron radiation point spectral area); **(b)** Spectra for the 3 area classes in the basalis regions of the three specimens; **(c)** Cost/ function plot identifying the optimal number of PCs to be used for PCA; **(d)** Scores plot graphically representing classification by PCA-LDA. The *x*-axis represents DF1 and the *y*-axis DF2; **(e)** Wavenumber selection for SPA-LDA; **(f)** Scores plot graphically representing classification by SPA-LDA. The *x*-axis represents DF1 and the *y*-axis DF2; **(g)** Wavenumber selection for GA-LDA; **(h)** Scores plot graphically representing classification by GA-LDA. The *x*-axis represents DF1 and the *y*-axis DF2.

Figure 5. Classification of epithelial cell classes furthest apart in the endometrial crypt base by synchrotron radiation spectral analysis using PCA-LDA, SPA-LDA and GA-LDA on FPA-FTIR microspectroscopy-derived data. **(a)** Micrograph with circles representing the regions sampled (this is a representation as the areas analysed were selected at different

magnifications to reveal the synchrotron point spectral area); **(b)** Pre-processed spectra of the classes furthest apart in crypt bases; **(c)** Cost/ function plot identifying the optimal number of PCs to be used for PCA; **(d)** Scores plot graphically representing classification by PCA-LDA. The *x*-axis represents Sample Index and the *y*-axis DF1; **(e)** Wavenumber selection for SPA-LDA; **(f)** Scores plot graphically representing classification by SPA-LDA. The *x*-axis represents Sample Index and the *y*-axis DF1; **(g)** Wavenumber selection for GA-LDA; **(h)** Scores plot graphically representing classification by GA-LDA. The *x*-axis represents Sample Index and the *y*-axis DF1.

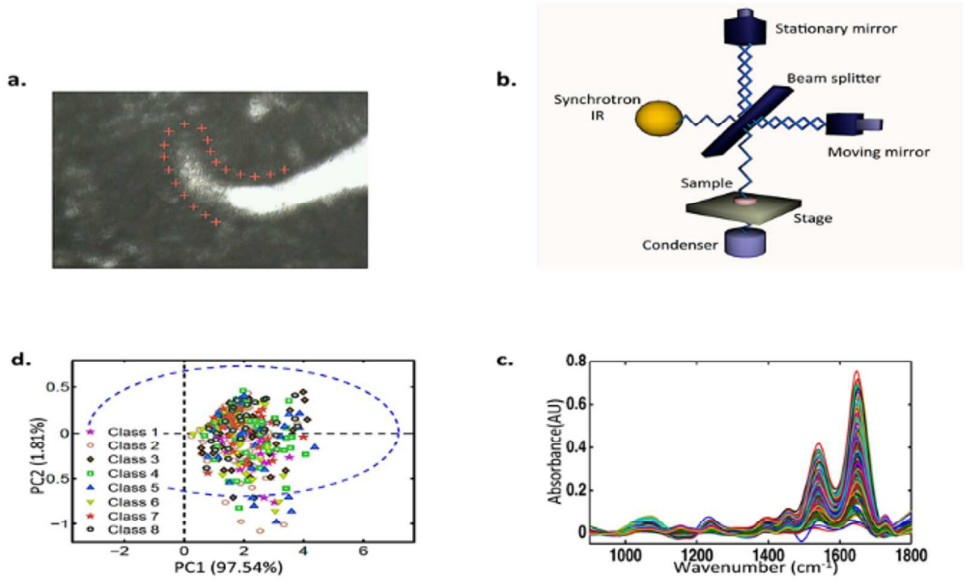
Figure 6. Exploratory analysis of synchrotron radiation-derived data from crypt bases to isolate putative stem cells and estimate their location. **(a)** Micrograph with circles representing the regions sampled (this is a representation as the regions analysed were selected at different magnifications to reveal the synchrotron radiation point spectral area); **(b)** Raw spectra of the classes (regions) sampled in crypt bases; **(c)** Pre-processed spectra of the classes (regions) sampled in crypt bases; **(d)** Scores plot graphically representing classification by PCA. The *x*-axis represents PC1 and the *y*-axis PC2, the dotted ellipse the 95 Confidence interval (CI); **(e)** Scores plot graphically representing classification by Q residuals. The horizontal dotted line represents 2 standard deviations; **(f)** Pre-processed spectra of the epithelial cells (blue) and "outliers" (green) sampled in crypt bases; **(g)** Averaged spectra of the epithelial cells (blue) and "outliers" (green) sampled in crypt bases; **(h)** Loadings curve to identify wavenumbers responsible for segregation of "outliers" from other epithelial cells. The peaks and troughs furthest from the dotted line are the most responsible wavenumber regions.

Figure 7. Intensity images based on FPA imaging (pink represents highest intensity while blue the lowest). **(a)** H&E-stained longitudinally cut uterine glandular crypt (circled in blue: cells in the deepest portion of the gland; circled in red: basalis/functionalis cells; circled in

734 black: cells close to the gland opening into the endometrial cavity); **(b)** White light image of
735 longitudinally-cut uterine glandular crypt (magnification 15×); **(c)** 2-D intensity map derived
736 following FPA spectral analysis at the wavenumber 1,080 cm^{-1} overlaid onto a white light
737 micrograph (magnification 15×); **(d)** 2-D intensity map derived following FPA spectral
738 analysis at the wavenumber 1,390 cm^{-1} overlaid onto a white light micrograph (magnification
739 15×); **(e)** 2-D intensity map derived following FPA spectral analysis at the wavenumber
740 1,550 cm^{-1} overlaid onto a white light micrograph (magnification 15×); and, **(f)** 2-D intensity
741 map derived following FPA spectral analysis at the wavenumber 1,650 cm^{-1} overlaid onto a
742 white light micrograph (magnification 15×). Colour bar: normalized absorbance for (c-f).

743

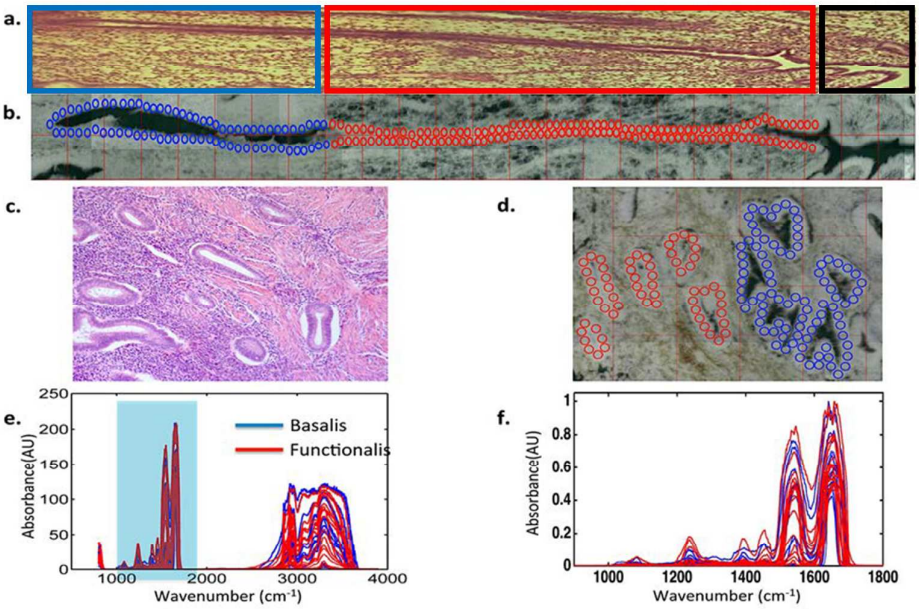
744 **Figure 1**



745

746

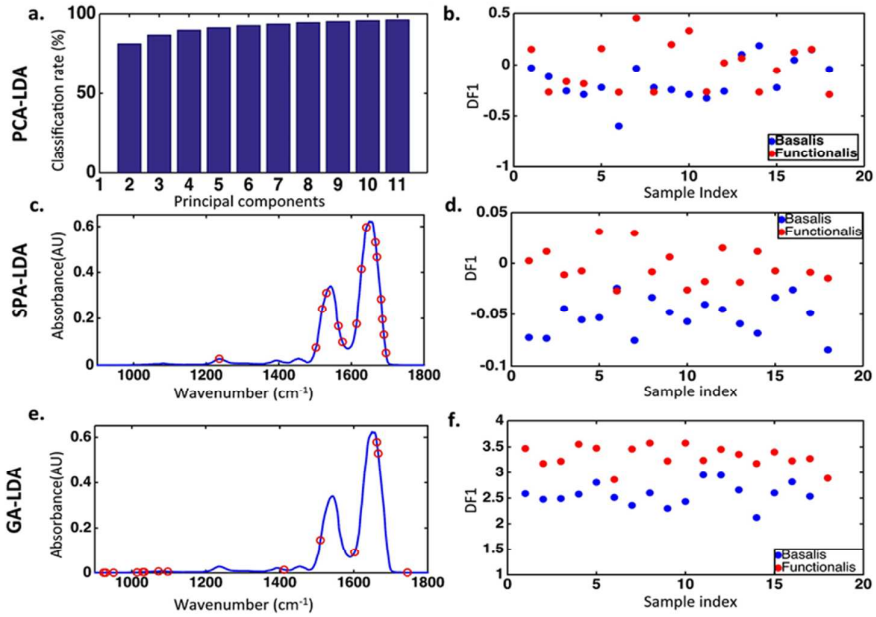
747 **Figure 2**



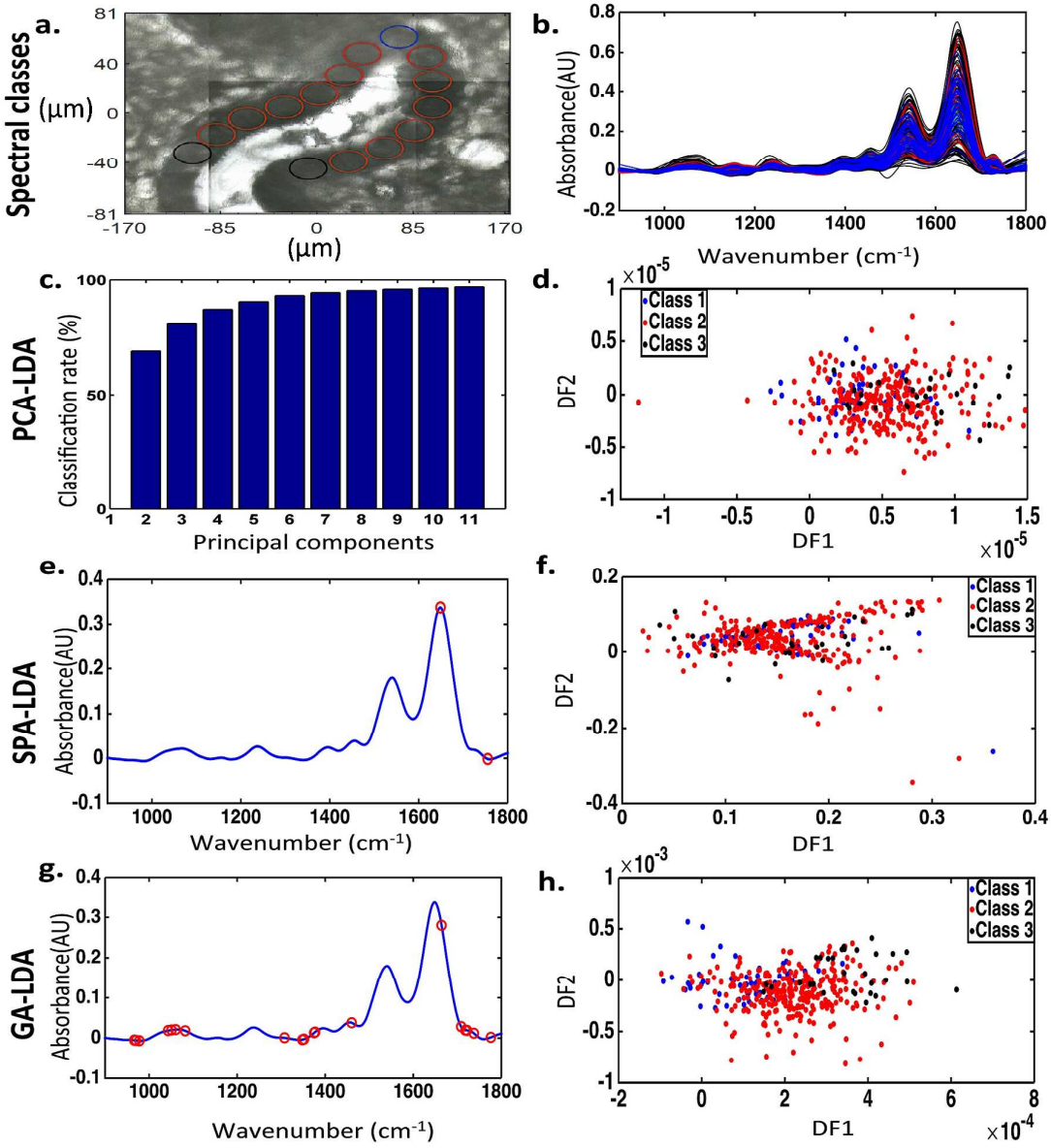
748

749

Figure 3



753 **Figure 4**



754

Figure 5

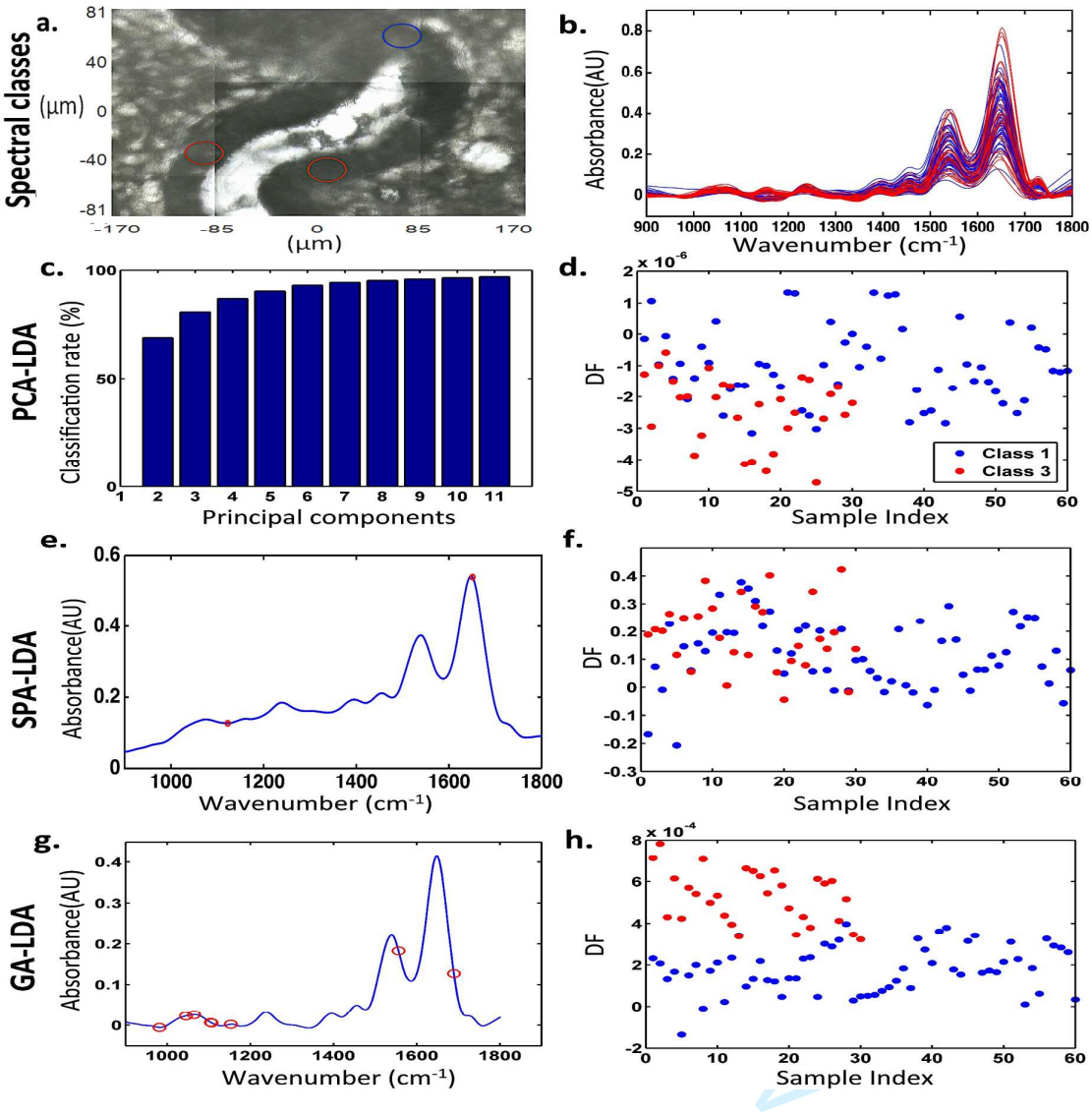
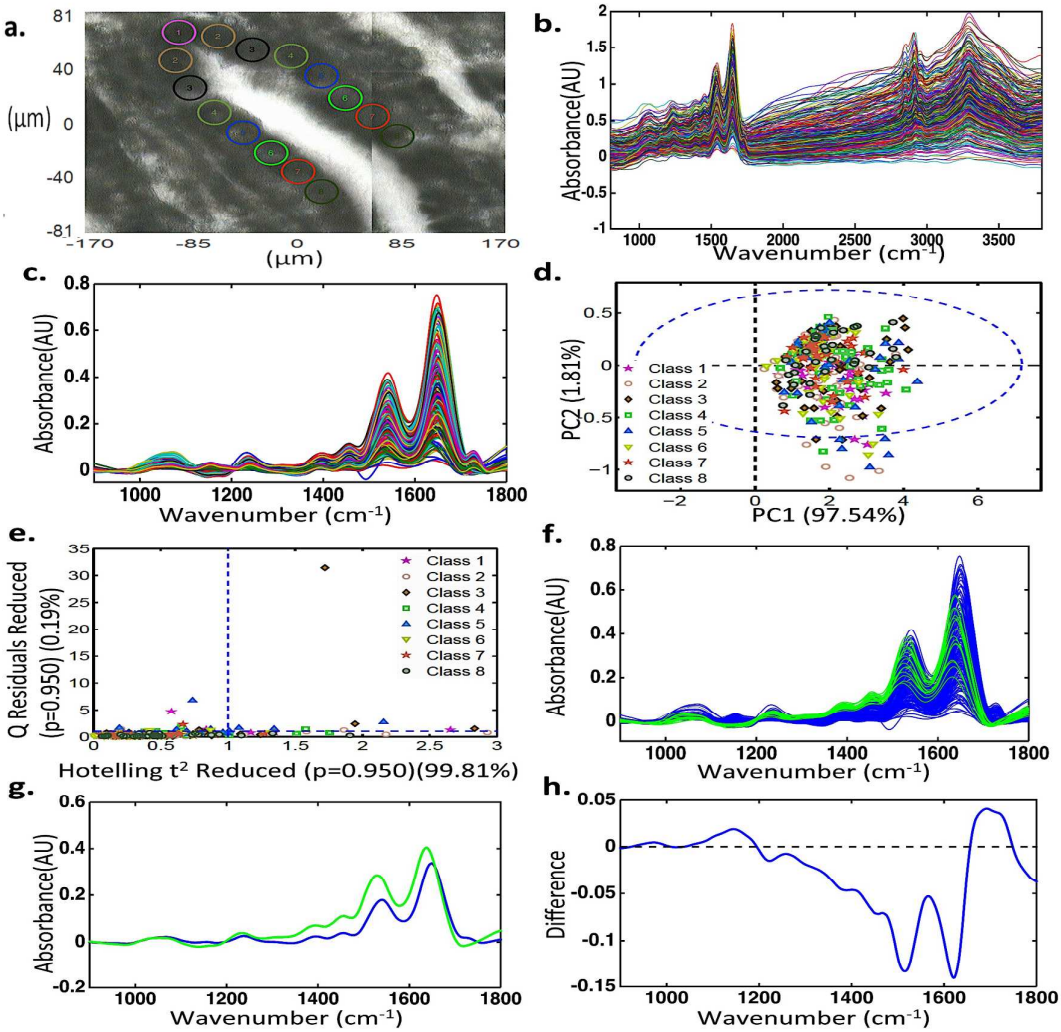
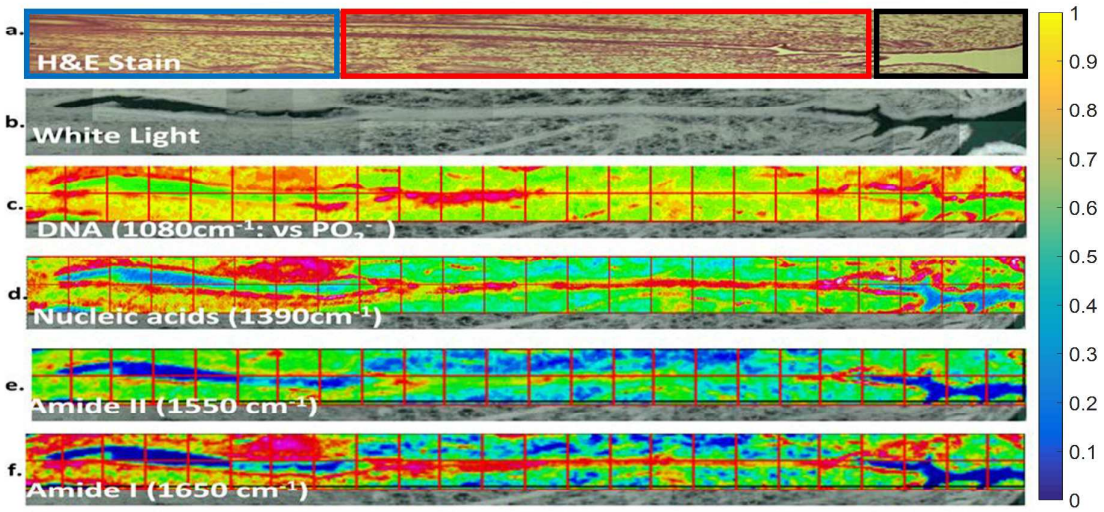

a.
Spectral classes
(μm)
(μm)
b.
Absorbance(AU)
Wavenumber (cm⁻¹)
c.
PCA-LDA
Classification rate (%)
Principal components
d.
DF
Class 1
Class 3
Sample Index
e.
SPA-LDA
Absorbance(AU)
Wavenumber (cm⁻¹)
f.
DF
Sample Index
g.
GA-LDA
Absorbance(AU)
Wavenumber (cm⁻¹)
h.
DF
Sample Index

Figure 6


a.
(μm)
(μm)
b.
Absorbance(AU)
Wavenumber (cm-1)
c.
Absorbance(AU)
Wavenumber (cm-1)
d.
PC2 (1.81%)
PC1 (97.54%)
Class 1
Class 2
Class 3
Class 4
Class 5
Class 6
Class 7
Class 8
e.
Q Residuals Reduced (p=0.950) (0.19%)
Hotelling t2 Reduced (p=0.950)(99.81%)
f.
Absorbance(AU)
Wavenumber (cm-1)
g.
Absorbance(AU)
Wavenumber (cm-1)
h.
Difference
Wavenumber (cm-1)

760 **Figure 7**



761

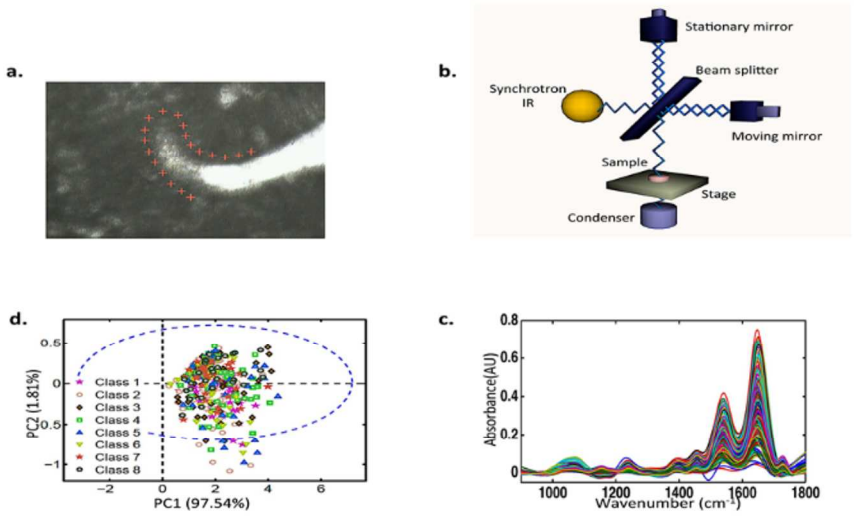


Figure 1. Synchrotron radiation-based FTIR spectral interrogation and analysis of endometrial epithelial gland bases to identify the location of putative stem cells. (a) Micrograph with superimposed marks representing areas interrogated by synchrotron radiation (magnification 36×); (b) Diagram representing a Michelson interferometer, which is the basic structure of an IR spectroscope; (c) The interferogram is transformed to an absorbance spectrum by Fourier-transform; and, (d) PCA identifies "outliers" which may represent putative stem cells.

338x190mm (96 x 96 DPI)

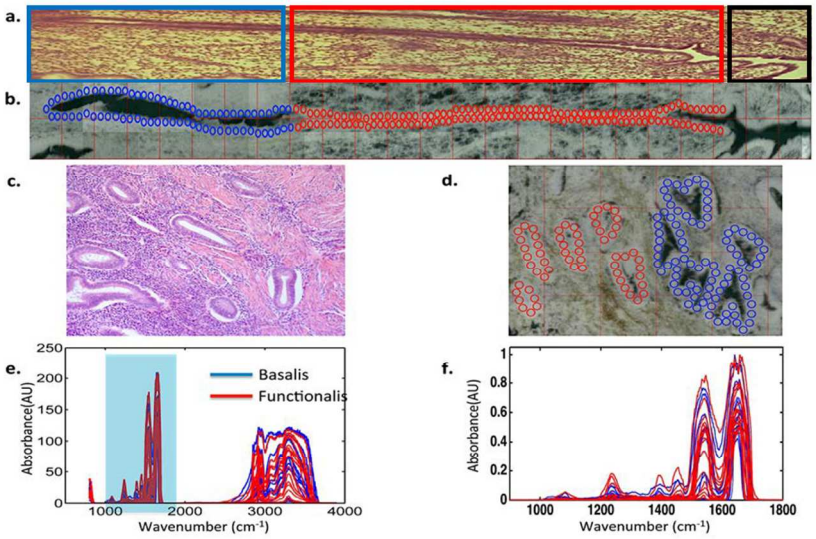


Figure 2. Representation of region identification for spectral selection by FPA and pre-processing of acquired spectra. (a) H&E-stained parallel section of examined longitudinally cut crypt, in the secretory phase (circled in blue: cells in the deepest portion of the gland; circled in red: basalis/functionalis cells; circled in black: cells close to the gland opening into the endometrial cavity); (b) Micrograph of examined longitudinally cut crypt with overlaid circular markers representing epithelial areas which were selected for comparison (magnification 15×); (the actual spectral area selection on the pseudo-image was performed at different magnifications to ensure the spectra selected were good representatives of the areas to be examined); (c) H&E-stained parallel section of examined coronally cut crypt, in the secretory phase; (d) Micrograph of examined coronally cut crypt with overlaid circular markers representing epithelial areas which were selected for comparison (magnification 15×) (the actual spectral area selection on the pseudo-image was performed at different magnifications to ensure the spectra selected were good representatives of the areas to be examined); (e) Averaged absorbance spectra acquired by FPA imaging of all samples; and, (f) Pre-processed spectra resulting from the raw spectra acquired by FPA imaging of all samples.

338x190mm (150 x 150 DPI)

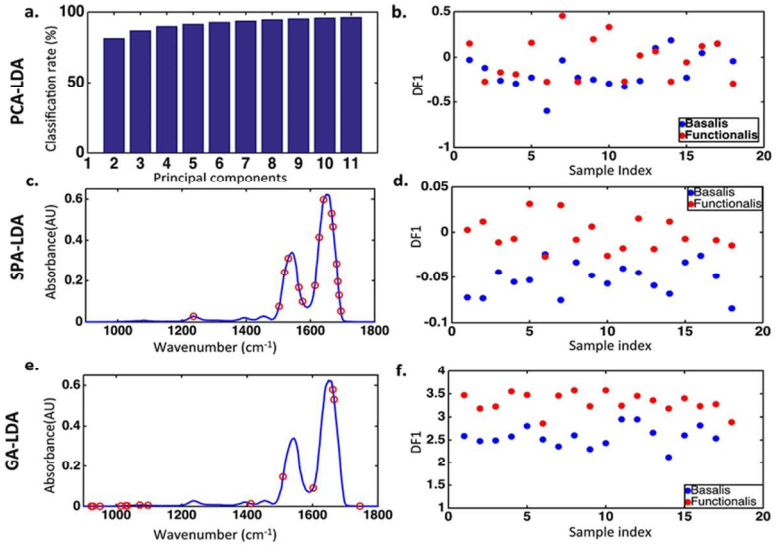


Figure 3. Classification of Basalis and Functionalis regions by spectral analysis using PCA-LDA, SPA-LDA and GA-LDA on FPA-FTIR derived data (red = Functionalis, blue = Basalis): All specimens. (a) Cost/function plot identifying the optimal number of PCs to be used for PCA; (b) Scores plot graphically representing classification by PCA-LDA. The x-axis represents the sample index and the y-axis the discriminant factor 1 (DF1); (c) Wavenumber selection for SPA-LDA; (d) Scores plot graphically representing classification by SPA-LDA. The x-axis represents the sample index and the y-axis DF1; (e) Wavenumber selection for GA-LDA; (f) Scores plot graphically representing classification by GA-LDA. The x-axis represents sample index and the y-axis DF1.

338x190mm (96 x 96 DPI)

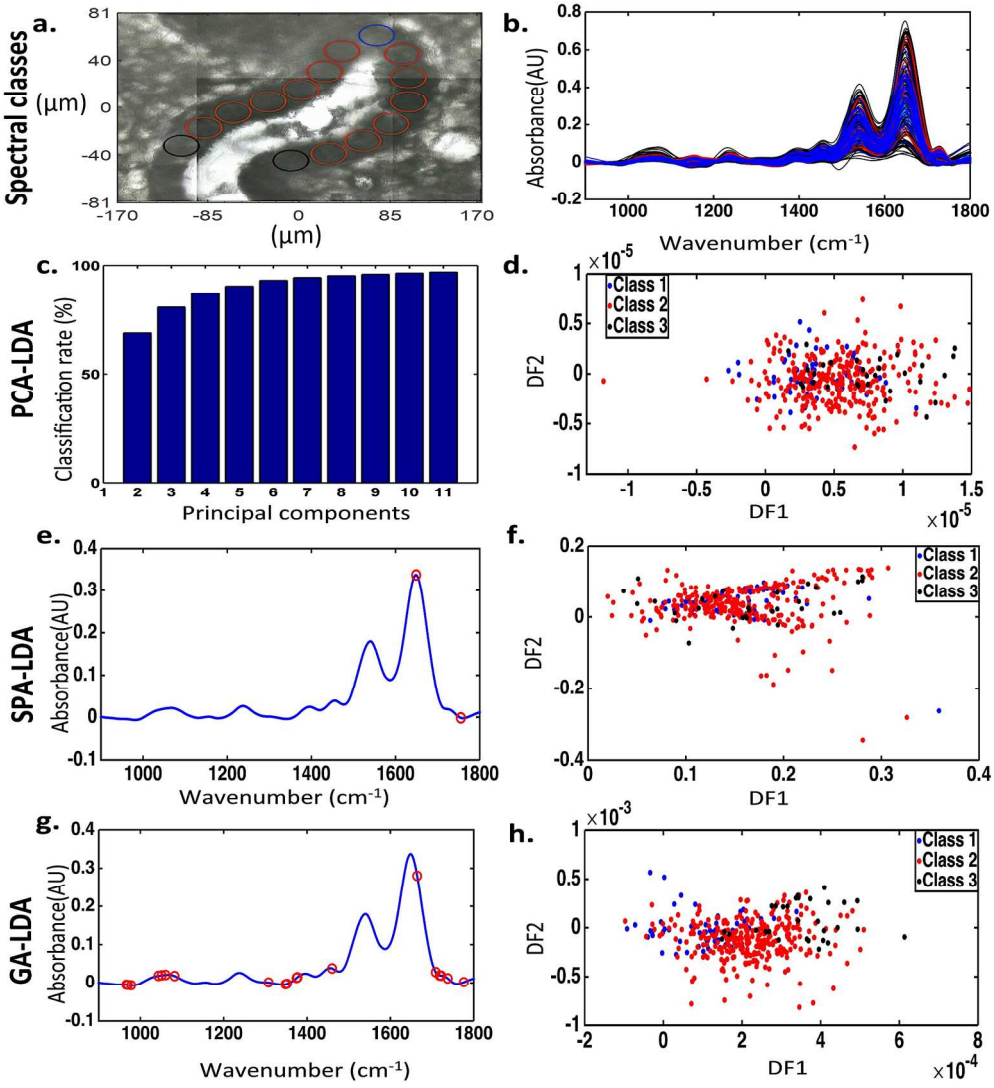


Figure 4. Classification of epithelial cells according to their position in the endometrial crypt base by synchrotron radiation spectral analysis using PCA-LDA, SPA-LDA and GA-LDA on FPA-FTIR microscopy-derived spectral data. (a) Micrograph with circles representing the regions sampled (this is a representation as the areas analysed were selected at different magnifications to reveal the synchrotron radiation point spectral area); (b) Spectra for the 3 area classes in the basalis regions of the three specimens; (c) Cost/function plot identifying the optimal number of PCs to be used for PCA; (d) Scores plot graphically representing classification by PCA-LDA. The x-axis represents DF1 and the y-axis DF2; (e) Wavenumber selection for SPA-LDA; (f) Scores plot graphically representing classification by SPA-LDA. The x-axis represents DF1 and the y-axis DF2; (g) Wavenumber selection for GA-LDA; (h) Scores plot graphically representing classification by GA-LDA. The x-axis represents DF1 and the y-axis DF2.

372x403mm (150 x 150 DPI)

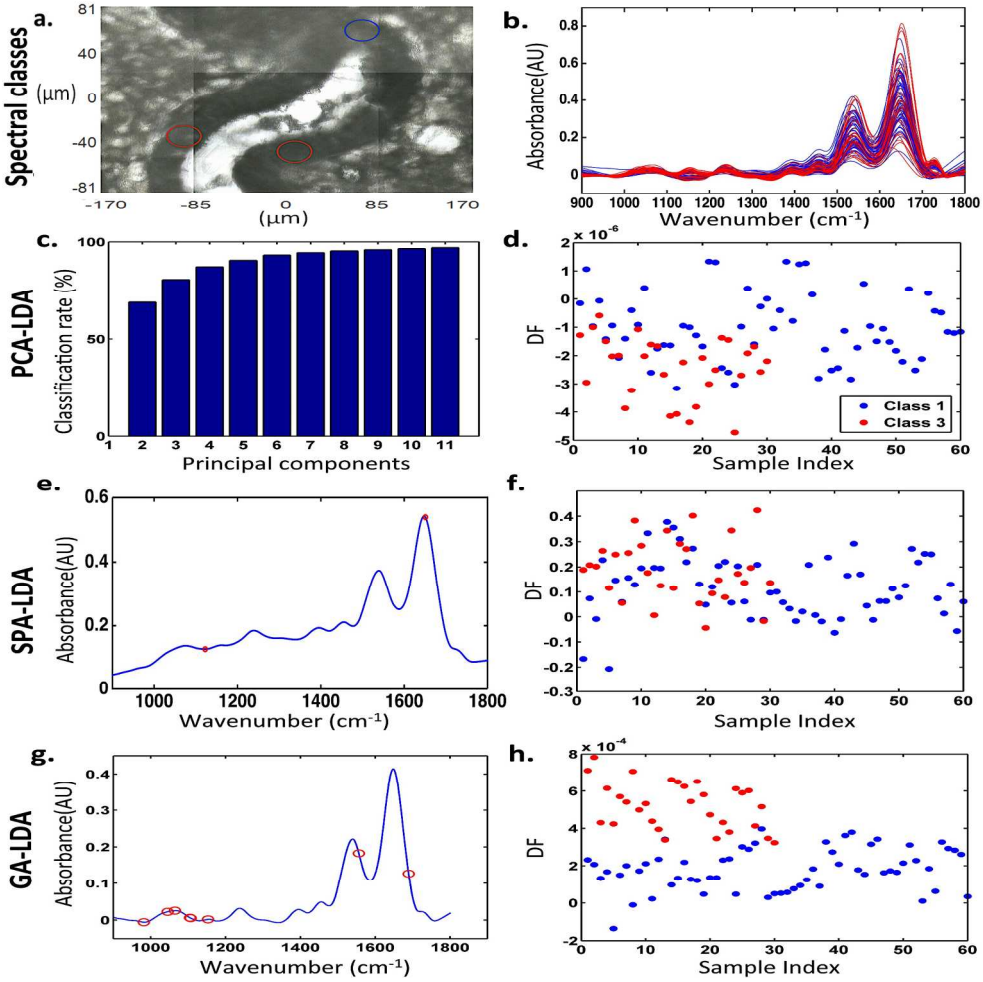


Figure 5. Classification of epithelial cell classes furthest apart in the endometrial crypt base by synchrotron radiation spectral analysis using PCA-LDA, SPA-LDA and GA-LDA on FPA-FTIR microspectroscopy-derived data. (a) Micrograph with circles representing the regions sampled (this is a representation as the areas analysed were selected at different magnifications to reveal the synchrotron point spectral area); (b) Pre-processed spectra of the classes furthest apart in crypt bases; (c) Cost/ function plot identifying the optimal number of PCs to be used for PCA; (d) Scores plot graphically representing classification by PCA-LDA. The x-axis represents Sample Index and the y-axis DF1; (e) Wavenumber selection for SPA-LDA; (f) Scores plot graphically representing classification by SPA-LDA. The x-axis represents Sample Index and the y-axis DF1; (g) Wavenumber selection for GA-LDA; (h) Scores plot graphically representing classification by GA-LDA. The x-axis represents Sample Index and the y-axis DF1.

411x403mm (150 x 150 DPI)

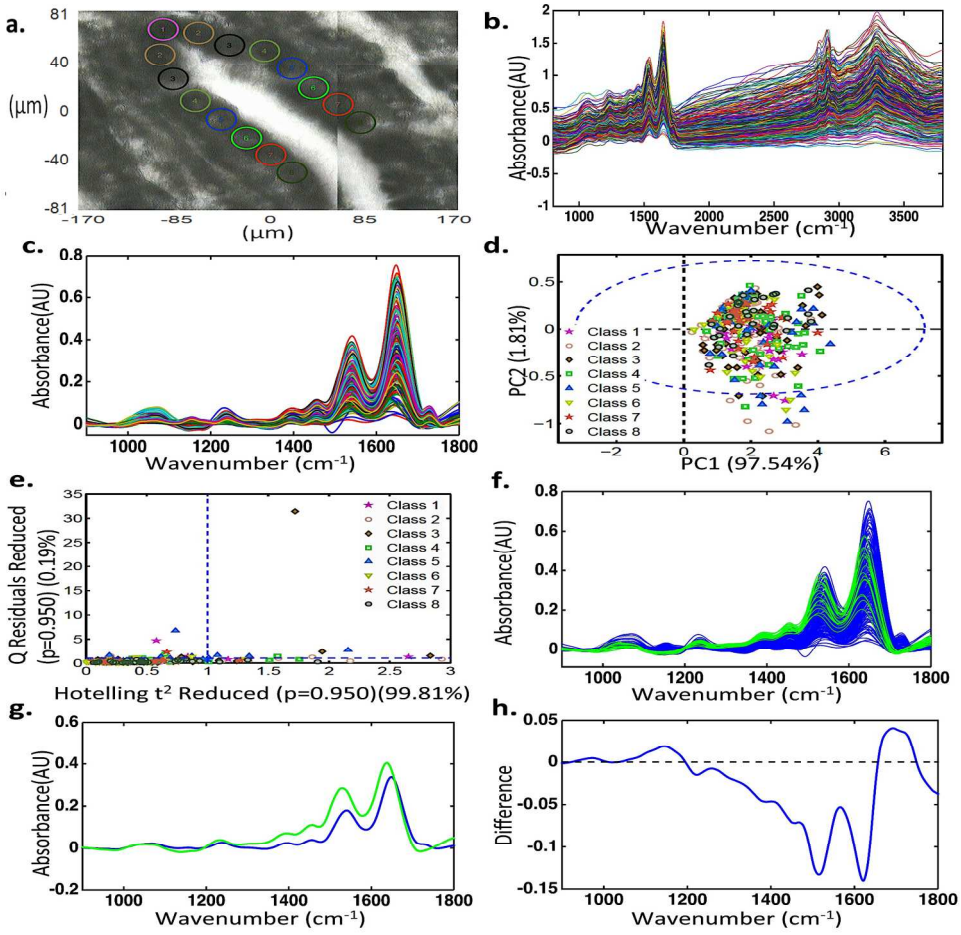


Figure 6. Exploratory analysis of synchrotron radiation-derived data from crypt bases to isolate putative stem cells and estimate their location. (a) Micrograph with circles representing the regions sampled (this is a representation as the regions analysed were selected at different magnifications to reveal the synchrotron radiation point spectral area); (b) Raw spectra of the classes (regions) sampled in crypt bases; (c) Pre-processed spectra of the classes (regions) sampled in crypt bases; (d) Scores plot graphically representing classification by PCA. The x-axis represents PC1 and the y-axis PC2, the dotted ellipse the 95 Confidence interval (CI); (e) Scores plot graphically representing classification by Q residuals. The horizontal dotted line represents 2 standard deviations; (f) Pre-processed spectra of the epithelial cells (blue) and "outliers" (green) sampled in crypt bases; (g) Averaged spectra of the epithelial cells (blue) and "outliers" (green) sampled in crypt bases; (h) Loadings curve to identify wavenumbers responsible for segregation of "outliers" from other epithelial cells. The peaks and troughs furthest from the dotted line are the most responsible wavenumber regions.

414x403mm (150 x 150 DPI)

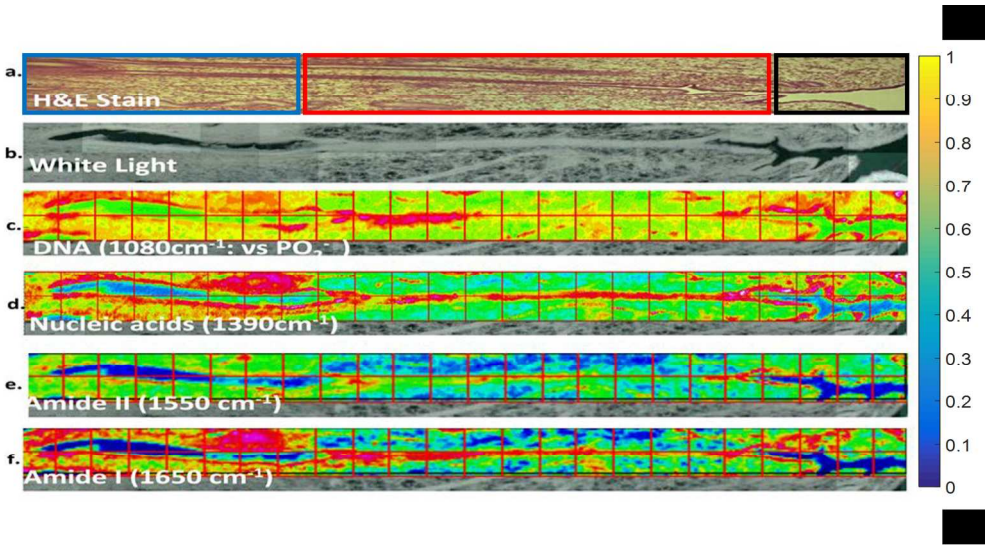


Figure 7. Intensity images based on FPA imaging (pink represents highest intensity while blue the lowest). (a) H&E-stained longitudinally cut uterine glandular crypt (circled in blue: cells in the deepest portion of the gland; circled in red: basalis/functionalis cells; circled in black: cells close to the gland opening into the endometrial cavity); (b) White light image of longitudinally-cut uterine glandular crypt (magnification 15×); (c) 2-D intensity map derived following FPA spectral analysis at the wavenumber 1,080 cm-1 overlaid onto a white light micrograph (magnification 15×); (d) 2-D intensity map derived following FPA spectral analysis at the wavenumber 1,390 cm-1 overlaid onto a white light micrograph (magnification 15×); (e) 2-D intensity map derived following FPA spectral analysis at the wavenumber 1,550 cm-1 overlaid onto a white light micrograph (magnification 15×); and, (f) 2-D intensity map derived following FPA spectral analysis at the wavenumber 1,650 cm-1 overlaid onto a white light micrograph (magnification 15×). Colour bar: normalized absorbance for (c-f).

353x190mm (150 x 150 DPI)

Electronic Supplementary Information

Synchrotron- and focal plane array-based Fourier-transform infrared spectroscopy differentiates the Basalis and Functionalis epithelial endometrial regions and identifies putative stem cell regions of human endometrial glands

Georgios Theophilou[1], Camilo L. M. Morais[2], Diane E Halliwell[2], Kássio M. G. Lima[3], Josephine Drury[4], Pierre L. Martin-Hirsch[5], Helen F. Stringfellow[5], Dharani K. Hapangama[4], Francis L. Martin[2*]

[1]Department of Gynaecology, Leeds Teaching Hospitals NHS Foundation Trust, Leeds, United Kingdom; [2]School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston PR1 2HE, United Kingdom; [3]Biological Chemistry and Chemometrics, Institute of Chemistry, Federal University of Rio Grande do Norte, Natal 59072-970, Brazil; [4]Department of Obstetrics and Gynaecology, Liverpool Women's NHS Foundation Trust, Liverpool, United Kingdom; [5]Department of Obstetrics and Gynaecology, Lancashire Teaching Hospitals NHS Foundation Trust, Preston, United Kingdom

***Corresponding author:** Francis L. Martin, School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston PR1 2HE, United Kingdom. Email: flmartin@uclan.ac.uk; Tel.: +44 01772 896482

Table S1. Number of averaged spectra arranged by area as selected on slides for FPA analysis.

Basalis		*Functionalis*	
Sampled area code	Number of averaged spectra	Sampled area code	Number of averaged spectra
1	2140	1	5127
2	510	2	945
3	1324	3	1534
4	2278	4	1589
5	3448	5	1534
6	4647	6	2064
7	1823	7	1299
8	1753	8	1100
9	1398	9	2060
10	1450	10	3876
11	1563	11	2155
12	803	12	1517
13	1906	13	3230
14	592	14	2525
15	1894	15	13211
16	803	16	10123
17	8195	17	7143
18	1830	18	4800
TOTAL	38357	TOTAL	65832

Table S2. Selected variables by each chemometric technique employed for the differentiation of *basalis* from *functionalis* epithelial regions using FPA spectral data.

	Selected variables as wavenumbers (cm^{-1})
SPA-LDA	1236, 1506, 1519, 1552, 1566, 1616, 1627, 1654, 1666, 1672, 1687, 1693
GA-LDA	923, 929, 950, 1014, 1029, 1033, 1072, 1097, 1411, 1510, 1602, 1662, 1666, 1746

Table S3. Overall classification rates by each chemometric technique employed to differentiate *basalis* from *functionalis* epithelial regions using FPA spectral data.

	PCA-LDA	SPA-LDA	GA-LDA
Classification rate for calibration set (%)	66.7	87.5	95.8
Classification rate for validation set (%)	50.0	100.0	100.0
Classification rate for prediction set (%)	33.3	100.0	100.0

Table S4. Overall classification rates by each chemometric technique employed to differentiate of *basalis* from *functionalis* tissues for individual specimen sets using FPA spectral data.

	PCA-LDA	SPA-LDA	GA-LDA
SPCN 1			
Classification rate for calibration set (%)	100.0	100.0	100.0
Classification rate for validation set (%)	66.7	100.0	100.0
Classification rate for prediction set (%)	100.0	75.0	50.0
SPCN 2			
Classification rate for calibration set (%)	100.0	100.0	100.0
Classification rate for validation set (%)	0.0	100.0	100.0
Classification rate for prediction set (%)	100.0	100.0	100.0
SPCN 3			
Classification rate for calibration set (%)	93.3	100.0	93.3
Classification rate for validation set (%)	100.0	100.0	100.0
Classification rate for prediction set (%)	100.0	66.7	66.7

Table S5. Selected variables by each chemometric technique employed to differentiate *basalis* from *functionalis* epithelial regions for individual specimen sets using FPA spectral data.

	Selected variables as wavenumbers (cm^{-1})
SPCN 1	
SPA-LDA	1496, 1554, 1567, 1602, 1618, 1627, 1662, 1681, 1685, 1691, 1695
GA-LDA	991, 1083, 1134, 1243, 1326, 1429, 1458, 1554, 1604, 1616, 1629, 1714
SPCN 2	
SPA-LDA	975, 1238, 1633, 1683
GA-LDA	973, 1251, 1616, 1620, 1658
SPCN 3	
SPA-LDA	1083, 1220, 1236, 1510, 1548, 1556, 1610, 1623, 1625, 1631, 1660, 1687, 1693
GA-LDA	933, 1139, 1297, 1299, 1747

Table S6. Overall classification rates for each chemometric technique employed to identify inter-individual variation between specimens within *basalis* epithelial layer.

	PCA-LDA	SPA-LDA	GA-LDA
Classification rate for calibration set (%)	93.3	100.0	100.0
Classification rate for validation set (%)	75.0	100.0	100.0
Classification rate for prediction set (%)	100.0	100.0	75.0

Table S7. Selected variables used for each chemometric technique employed to identify inter-individual variability within the *basalis* layer using FPA spectral data.

	Selected variables as wavenumbers (cm^{-1})
SPA-LDA	1236, 1548, 1564, 1585, 1602, 1614, 1633, 1648, 1666, 1683, 1689, 1743
GA-LDA	900, 983, 991, 1016, 1091, 1122, 1137, 1182, 1203, 1681, 1753

Table S8. Selected variables used for each chemometric technique employed to differentiate Class 1 and Class 3 regions at the epithelial crypt bases using synchrotron data.

	Selected variables as wavenumbers (cm^{-1})
SPA-LDA	1265, 1647
GA-LDA	981, 1045, 1064, 1105, 1107, 1153, 1556, 1689

Table S9. Overall classification rates by each chemometric technique employed to differentiate Class 1 and Class 3 regions at the epithelial crypt bases using synchrotron data.

	PCA-LDA	SPA-LDA	GA-LDA
Classification rate for calibration set (%)	74.6	58.7	96.8
Classification rate for validation set (%)	69.2	69.2	100.0
Classification rate for prediction set (%)	78.6	50.0	92.9

Figure S1. Loadings curves derived from PCA-LDA for the comparison of basalis and functionalis epithelial regions using FPA spectroscopy; (Black line: Loadings on PC1; green line: Loadings on PC). **a.** Loadings curve for *basalis* and *functionalis* comparison for all specimens, **b.** Loadings curve for *basalis* and *functionalis* comparison for SPCN 1, **c.** Loadings curve for *basalis* and *functionalis* comparison for SPCN 2, **d.** Loadings curve for *basalis* and *functionalis* comparison for SPCN 3.

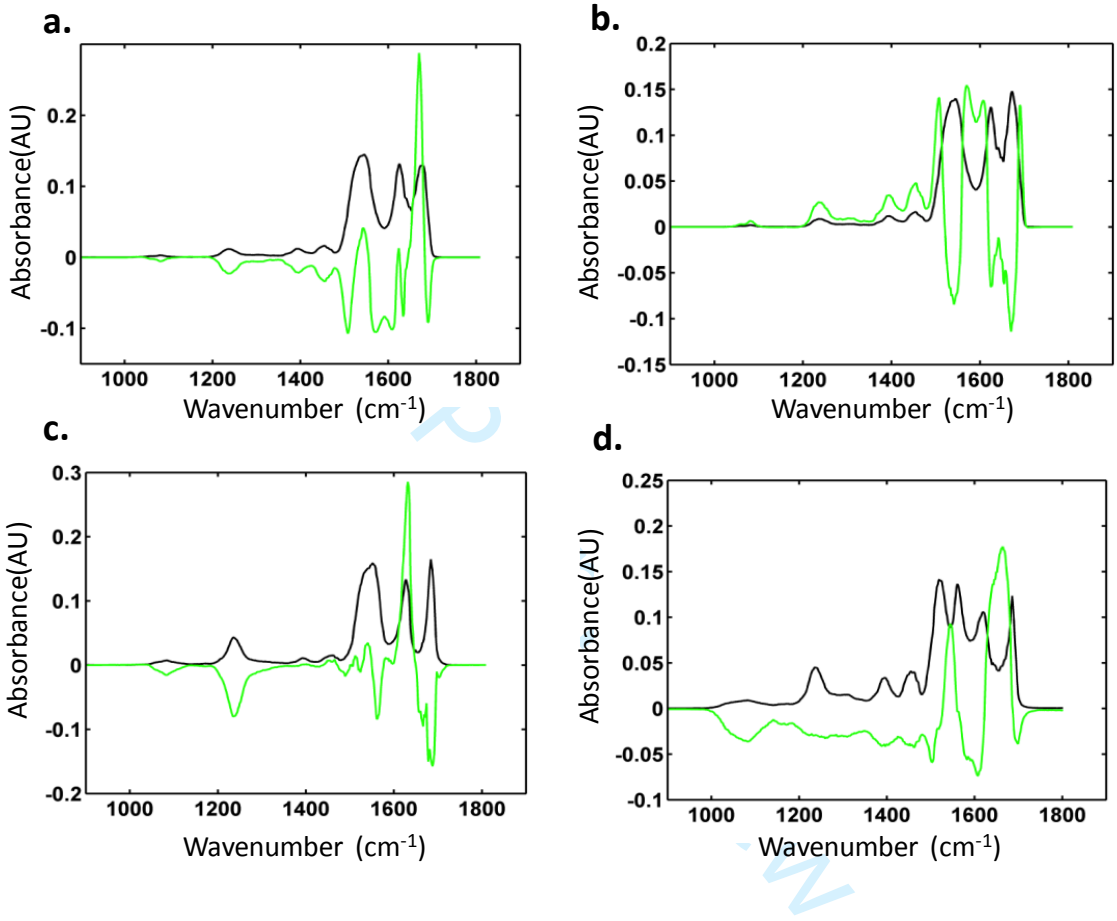


Figure S2. Classification of *Basalis* and *Functionalis* regions by spectral analysis using PCA-LDA, SPA-LDA and GA-LDA on FPA-FTIR derived data. (red= *Functionalis*, blue= *Basalis*): SPCN 1. **a.** Pre-processed spectra containing the biological fingerprint spectral regions, b. Averaged spectra for the *basalis* and *functionalis* regions, **c.** Cost/ function plot identifying the optimal number of PCs to be used for PCA, **d.** Scores plot graphically representing classification by PCA-LDA. The *x*-axis represents the sample index and the *y*-axis DF1, **e.** Wavenumber selection for SPA-LDA, **f.** Scores plot graphically representing classification by SPA-LDA. The *x*-axis represents the sample index and the *y*-axis DF1, **g.** Wavenumber selection for GA-LDA, **h.** Scores plot graphically representing classification by GA-LDA. The *x*-axis represents sample index and the *y*-axis DF1.

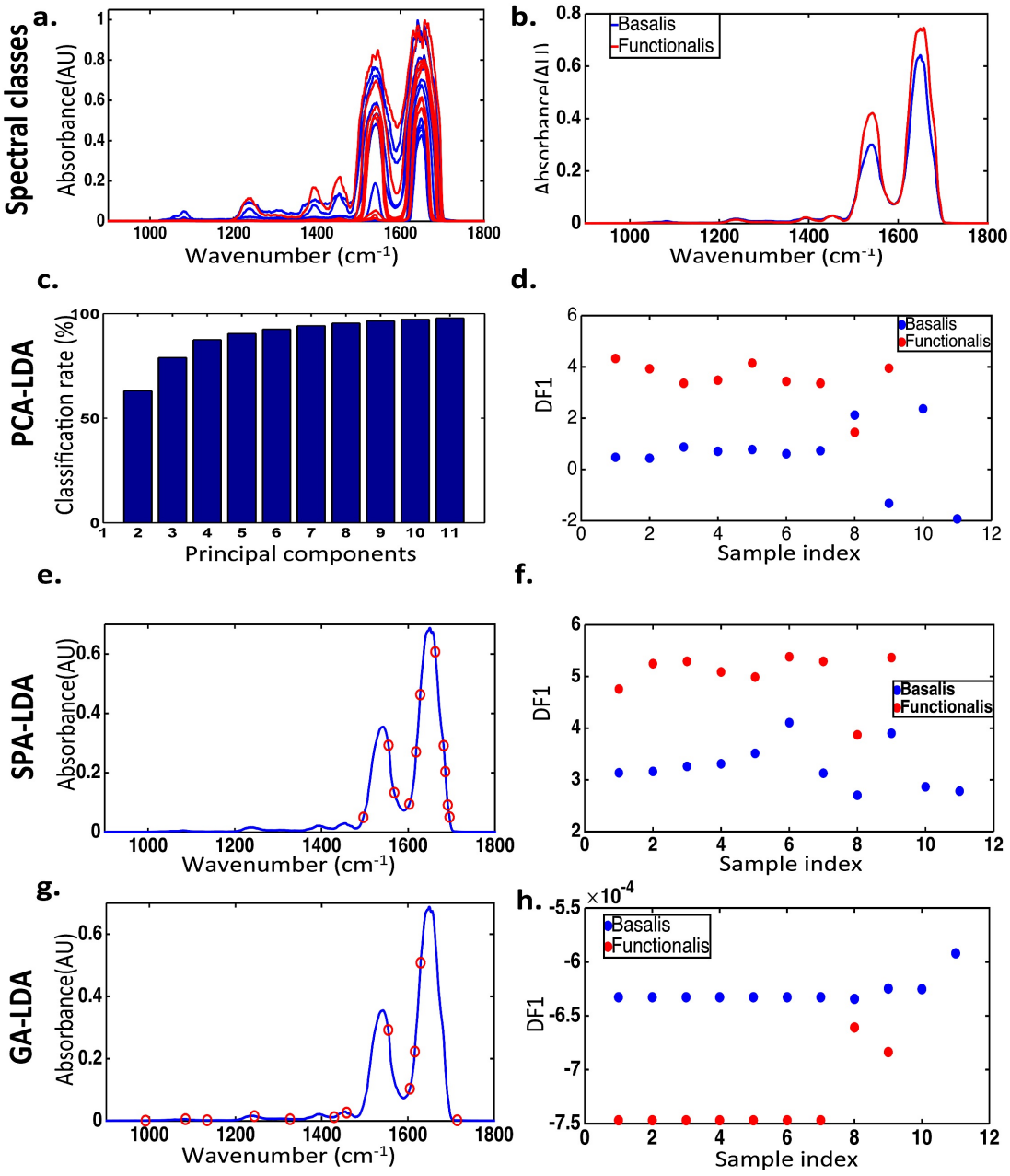


Figure S3. **Classification of *Basalis* and *Functionalis* regions by spectral analysis using PCA-LDA, SPA-LDA and GA-LDA on FPA-FTIR derived data. (red= *Functionalis*, blue= *Basalis*): SPCN2**. **a.** Pre-processed spectra containing the biological fingerprint spectral regions, **b.** Averaged spectra for the *basalis* and *functionalis* regions, **c.** Cost/ function plot identifying the optimal number of PCs to be used for PCA, **d.** Scores plot graphically representing classification by PCA-LDA. The *x*-axis represents the sample index and the *y*-axis DF1, **e.** Wavenumber selection for SPA-LDA, **f.** Scores plot graphically representing classification by SPA-LDA. The *x*-axis represents the sample index and the *y*-axis DF1, **g.** Wavenumber selection for GA-LDA, **h.** Scores plot graphically representing classification by GA-LDA. The *x*-axis represents sample index and the *y*-axis DF1.

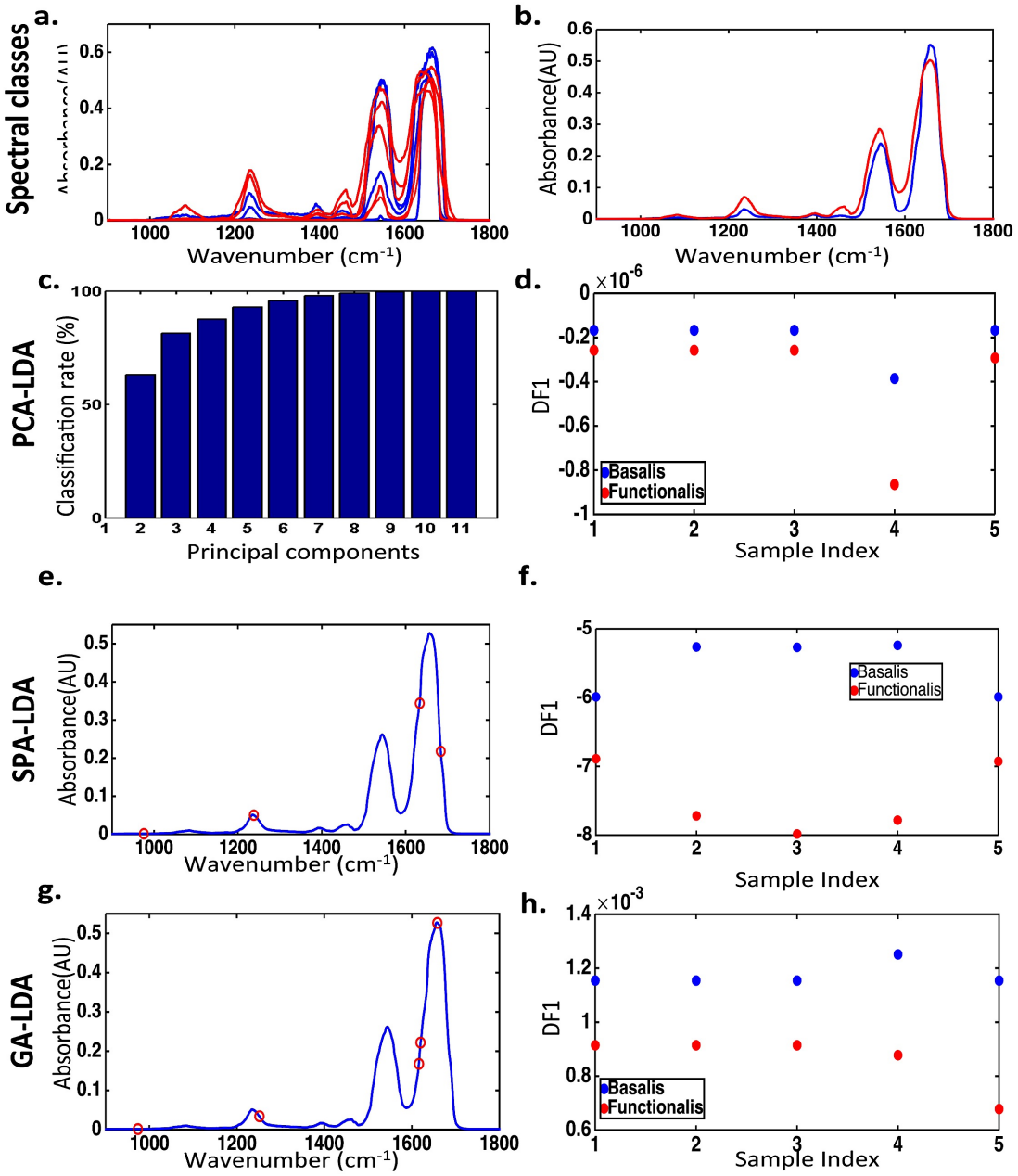


Figure S4. Classification of *Basalis* and *Functionalis* regions by spectral analysis using PCA-LDA, SPA-LDA and GA-LDA on FPA-FTIR derived data. (red= *Functionalis*, blue= *Basalis*): SPCN 3. **a.** Pre-processed spectra containing the biological fingerprint spectral regions, **b.** Averaged spectra for the *basalis* and *functionalis* regions, **c.** Cost/ function plot identifying the optimal number of PCs to be used for PCA, **d.** Scores plot graphically representing classification by PCA-LDA. The *x*-axis represents the sample index and the *y*-axis DF1, **e.** Wavenumber selection for SPA-LDA, **f.** Scores plot graphically representing classification by SPA-LDA. The *x*-axis represents the sample index and the *y*-axis DF1, **g.** Wavenumber selection for GA-LDA, **h.** Scores plot graphically representing classification by GA-LDA. The *x*-axis represents sample index and the *y*-axis DF1.

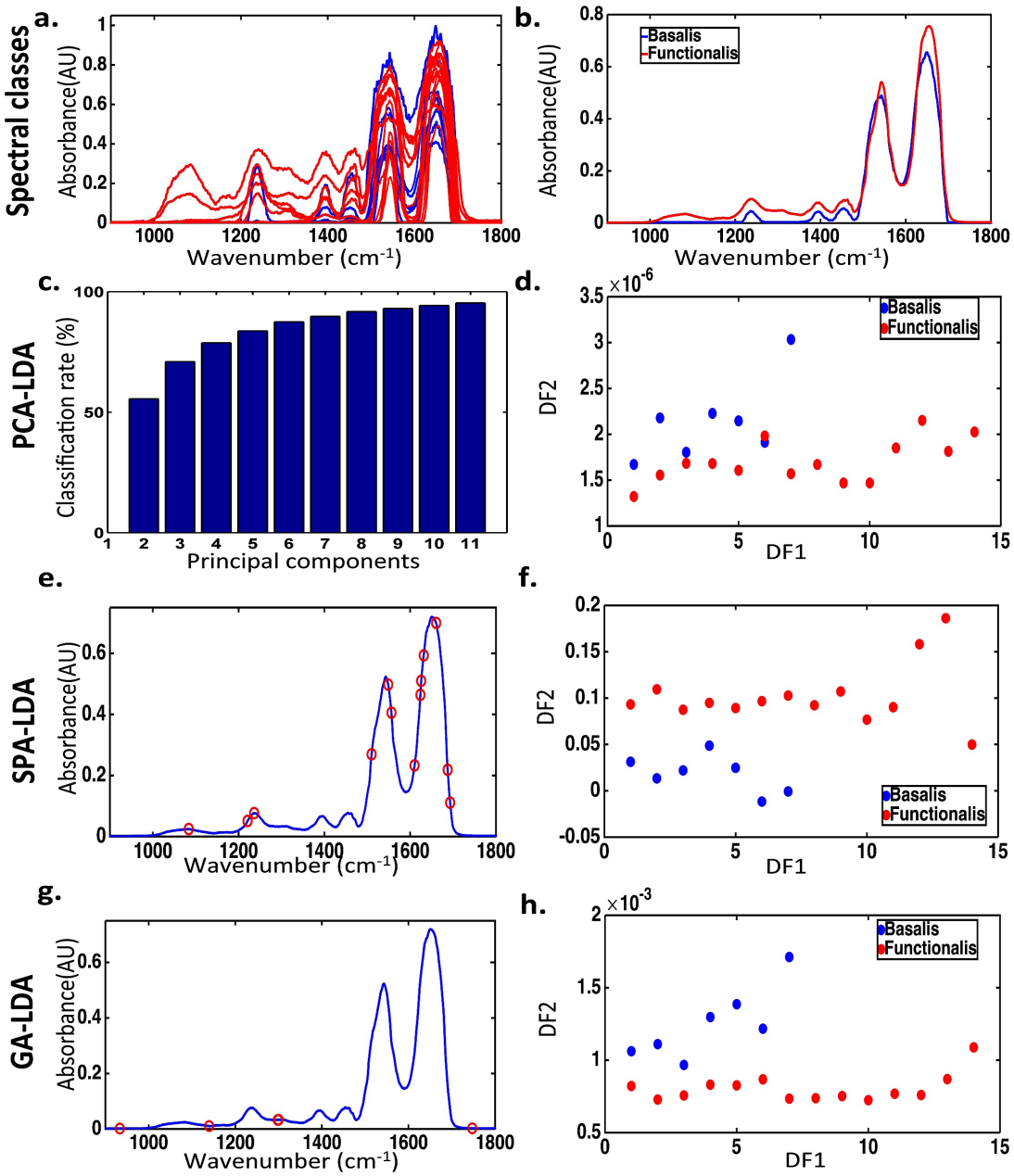


Figure S5. Classification specimens based on variability of their *Basalis* regions by spectral analysis using PCA-LDA, SPA-LDA and GA-LDA on FPA-FTIR derived data. (blue= SPCN1, red= SPCN2, black, SPCN3). **a.** Pre-processed spectra containing the biological fingerprint spectral regions, **b.** Averaged spectra for the *basalis* regions of the three specimens, **c.** Cost/ function plot identifying the optimal number of PCs to be used for PCA, **d.** Scores plot graphically representing classification by PCA-LDA. The *x*-axis represents DF1 and the *y*-axis DF2, **e.** Wavenumber selection for SPA-LDA, f. Scores plot graphically representing classification by SPA-LDA. The *x*-axis represents DF1 and the *y*-axis DF2, **g.** Wavenumber selection for GA-LDA, **h.** Scores plot graphically representing classification by GA-LDA. The *x*-axis represents DF1 and the *y*-axis DF2.